STIC-Biotech/Ch mLib 71354

From: Chan, Christina
S nt: Thursday, July 18, 2002 4:21 PM
To: Sullivan, Daniel; STIC-Biotech/ChemLib
Subject: RE: CDB Search Request for 2 month amended

Pl ase rush. Thanks Chris

-----Original Message-----
From: Sullivan, Daniel
Sent: Thursday, July 18, 2002 4:06 PM
T : Chan, Christina
Subject: FW: CDB Search Request for 2 month amended
Imp rtance: High

Hi Chris,

Could you please approve this search request for me? Thanks.

Dan

-----Original Message-----
From: Sullivan, Daniel
Sent: Thursday, July 18, 2002 3:55 PM
T : STIC-Biotech/ChemLib
Subject: CDB Search Request for 2 month amended
Imp rtance: High

Please do a **RUSH** search for two month amended case **09/754014**, nucleic acids 1-9, 16-22, and 25-45 of SEQ ID NO: 10 against the commercial or interference nucleic acid databases or both. If possible, the search can be limited to noncoding sequences within plasmids.

Thanks very much.

Daniel M. Sullivan
Examiner AU 1636
Room: 12D12
Mail Box: 11E12
Tel: 703-305-4448

Point of Contact:
Beverly Shears
Technical Info. Specialist
CM1 1E05 Tel: 308-4994

RECEIVED
JUL 18 2002
(STIC)

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: _____
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

U.S. DEPARTMENT OF COMMERCE
Patent and Trademark Office

SEARCH REQUEST FORM

Requestor's Name: _____ Serial Number: _____

Date: _____ Phone: _____ Art Unit: _____

Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevent citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevent claim(s).

STAFF USE ONLY

Date completed: 07-22-02
Searcher: Beverly C 4994
Terminal time: 20
Elapsed time: _____
CPU time: _____
Total time: 25
Number of Searches: _____
Number of Databases: 1

Search Site

_____ STIC
_____ CM-1
_____ Pre-S

Type of Search

_____ N.A. Sequence
_____ A.A. Sequence
_____ Structure
_____ Bibliographic

Vendors

_____ IG
_____ STN
_____ Dialog
_____ APS
_____ Geninfo
_____ SDC
_____ DARC/Questel
✓ Other CGN

PTO-1590 (9-90)

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2002, 09:45:13 ; Search time 2038.31 Seconds
(without alignments)
71.866 Million cell updates/sec

Title: US-09-754-014-10_COPY_16_22
Perfect score: 7
Sequence: 1 TACTAAC 7

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		7	100.0	7	6	AX203149	AX203149 Sequence
2	c	7	100.0	11	6	AX175038	AX175038 Sequence
3		7	100.0	11	6	AX175039	AX175039 Sequence
4		7	100.0	12	6	AR121253	AR121253 Sequence
5	c	7	100.0	13	6	AR037996	AR037996 Sequence
6	c	7	100.0	13	6	AR039094	AR039094 Sequence
7	c	7	100.0	13	6	AR050339	AR050339 Sequence
8	c	7	100.0	13	6	AR062898	AR062898 Sequence
9		7	100.0	14	6	I76088	I76088 Sequence 4
10		7	100.0	17	6	AR039499	AR039499 Sequence
11		7	100.0	17	6	AR039501	AR039501 Sequence
12		7	100.0	17	6	AR039503	AR039503 Sequence
13		7	100.0	17	6	AR040467	AR040467 Sequence
14		7	100.0	17	6	AR040469	AR040469 Sequence
15		7	100.0	17	6	AR040471	AR040471 Sequence
16		7	100.0	17	6	AR040473	AR040473 Sequence
17	c	7	100.0	17	6	AX214703	AX214703 Sequence
18	c	7	100.0	17	6	AX214704	AX214704 Sequence
19	c	7	100.0	17	6	AX214705	AX214705 Sequence
20	c	7	100.0	17	6	AX214706	AX214706 Sequence
21	c	7	100.0	17	6	AX214707	AX214707 Sequence
22	c	7	100.0	17	6	AX216434	AX216434 Sequence
23	c	7	100.0	17	6	AX216435	AX216435 Sequence
24	c	7	100.0	17	6	AX216696	AX216696 Sequence
25		7	100.0	17	6	AX217512	AX217512 Sequence
26		7	100.0	17	6	AX217513	AX217513 Sequence
27		7	100.0	17	6	AX217514	AX217514 Sequence
28		7	100.0	17	6	AX217873	AX217873 Sequence
29		7	100.0	17	6	AX217874	AX217874 Sequence
30		7	100.0	17	6	AX217875	AX217875 Sequence
31		7	100.0	17	6	AX217876	AX217876 Sequence
32		7	100.0	17	6	AX217877	AX217877 Sequence
33		7	100.0	17	6	AX218189	AX218189 Sequence
34		7	100.0	17	6	AX325433	AX325433 Sequence
35	c	7	100.0	17	6	AX325434	AX325434 Sequence
36		7	100.0	17	6	E01964	E01964 DNA encodin
37	c	7	100.0	17	6	E03611	E03611 DNA primer
38		7	100.0	17	6	I26214	I26214 Sequence 3
39	c	7	100.0	18	6	A68284	A68284 Sequence 5
40	c	7	100.0	18	6	AR062161	AR062161 Sequence
41		7	100.0	18	6	AR083039	AR083039 Sequence
42	c	7	100.0	18	6	AR108111	AR108111 Sequence
43		7	100.0	18	6	AR124377	AR124377 Sequence
44	c	7	100.0	18	6	AX003354	AX003354 Sequence
45	c	7	100.0	19	6	A22790	A22790 NCA HCV pro

ALIGNMENTS

RESULT 1
AX203149
LOCUS AX203149 7 bp mRNA linear PAT 30-AUG-2001
DEFINITION Sequence 2 from Patent WO0153529.
ACCESSION AX203149
VERSION AX203149.1 GI:15392500
KEYWORDS .
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequence.
REFERENCE 1 (bases 1 to 7)
AUTHORS Thomann,H.U. and Fitzgerald,M.S.
TITLE Rapid determination of gene structure using cdna sequence
JOURNAL Patent: WO 0153529-A 2 26-JUL-2001;
Genome Therapeutics Corporation (US)
FEATURES Location/Qualifiers
source 1. .7
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Yeast consensus branch site"
BASE COUNT 3 a 2 c 0 g 2 t
ORIGIN

```
  Query Match            100.0%;  Score 7;  DB 6;  Length 7;
  Best Local Similarity  100.0%;  Pred. No. 3e+09;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TACTAAC 7
          |||||||
Db      1 TACTAAC 7


RESULT   2
AX175038/c
LOCUS       AX175038                 11 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 27 from Patent WO0142493.
ACCESSION   AX175038
VERSION     AX175038.2  GI:15142057
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 11)
  AUTHORS   Olek,A. and Piepenbrock,C.
  TITLE     Method for the parallel detection of the degree of methylation of
            genomic dna
  JOURNAL   Patent: WO 0142493-A 27 14-JUN-2001;
            Epigenomics AG (DE)
COMMENT     On Aug 9, 2001 this sequence version replaced gi:14598498.
FEATURES             Location/Qualifiers
     source          1. .11
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="chemisch vorbehandelte Genom-DNA"
BASE COUNT        3 a      0 c      4 g      4 t
ORIGIN


  Query Match            100.0%;  Score 7;  DB 6;  Length 11;
  Best Local Similarity  100.0%;  Pred. No. 1.7e+06;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TACTAAC 7
          |||||||
Db     11 TACTAAC 5


RESULT   3
AX175039
LOCUS       AX175039                 11 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 28 from Patent WO0142493.
ACCESSION   AX175039
VERSION     AX175039.2  GI:15142058
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 11)
  AUTHORS   Olek,A. and Piepenbrock,C.
  TITLE     Method for the parallel detection of the degree of methylation of
            genomic dna
  JOURNAL   Patent: WO 0142493-A 28 14-JUN-2001;
            Epigenomics AG (DE)
COMMENT     On Aug 9, 2001 this sequence version replaced gi:14598499.
FEATURES             Location/Qualifiers
     source          1. .11
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="chemisch vorbehandelte Genom-DNA"
BASE COUNT        4 a      4 c      0 g      3 t
ORIGIN


  Query Match            100.0%;  Score 7;  DB 6;  Length 11;
  Best Local Similarity  100.0%;  Pred. No. 1.7e+06;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TACTAAC 7
          |||||||
Db      1 TACTAAC 7


RESULT   4
AR121253
LOCUS       AR121253                 12 bp    DNA     linear   PAT 16-MAY-2001
DEFINITION  Sequence 33 from patent US 6159710.
ACCESSION   AR121253
VERSION     AR121253.1  GI:14104829
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 12)
  AUTHORS   Fraser,N.W., Zabolotny,J.M. and Krummenacher,C.F.
  TITLE     Method and compositions for stabilizing unstable gene transcripts
  JOURNAL   Patent: US 6159710-A 33 12-DEC-2000;
FEATURES             Location/Qualifiers
     source          1. .12
                     /organism="unknown"
BASE COUNT        4 a      3 c      2 g      3 t
ORIGIN


  Query Match            100.0%;  Score 7;  DB 6;  Length 12;
  Best Local Similarity  100.0%;  Pred. No. 1.6e+06;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TACTAAC 7
          |||||||
Db      6 TACTAAC 12


RESULT   5
AR037996/c
LOCUS       AR037996                 13 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 2 from patent US 5804407.
ACCESSION   AR037996
VERSION     AR037996.1  GI:5956713
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 13)
  AUTHORS   Tamaoki,T. and Nakabayashi,H.
  TITLE     Method of expressing genes in mammalian cells
  JOURNAL   Patent: US 5804407-A 2 08-SEP-1998;
FEATURES             Location/Qualifiers
     source          1. .13
                     /organism="unknown"
BASE COUNT        4 a      1 c      2 g      6 t
ORIGIN


  Query Match            100.0%;  Score 7;  DB 6;  Length 13;
  Best Local Similarity  100.0%;  Pred. No. 1.6e+06;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TACTAAC 7
          |||||||
Db      7 TACTAAC 1


RESULT   6
AR039094/c
LOCUS       AR039094                 13 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 2 from patent US 5807738.
ACCESSION   AR039094
VERSION     AR039094.1  GI:5958457
```

```
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 13)
  AUTHORS   Tamaoki,T. and Nakabayashi,H.
  TITLE     Method of expressing genes in mammalian cells
  JOURNAL   Patent: US 5807738-A 2 15-SEP-1998;
FEATURES             Location/Qualifiers
     source          1. .13
                     /organism="unknown"
BASE COUNT        4 a      1 c      2 g      6 t
ORIGIN


  Query Match             100.0%;  Score 7;  DB 6;  Length 13;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+06;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TACTAAC 7
          |||||||
Db      7 TACTAAC 1


RESULT   7
AR050339/c
LOCUS       AR050339                13 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 2 from patent US 5827686.
ACCESSION   AR050339
VERSION     AR050339.1  GI:5973064
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 13)
  AUTHORS   Tamaoki,T. and Nakabayashi,H.
  TITLE     Method of expressing genes in mammalian cells
  JOURNAL   Patent: US 5827686-A 2 27-OCT-1998;
FEATURES             Location/Qualifiers
     source          1. .13
                     /organism="unknown"
BASE COUNT        4 a      1 c      2 g      6 t
ORIGIN


  Query Match             100.0%;  Score 7;  DB 6;  Length 13;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+06;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TACTAAC 7
          |||||||
Db      7 TACTAAC 1


RESULT   8
AR062898/c
LOCUS       AR062898                13 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 2 from patent US 5843776.
ACCESSION   AR062898
VERSION     AR062898.1  GI:5990589
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 13)
  AUTHORS   Tamaoki,T. and Nakabayashi,H.
  TITLE     Method of expressing genes in mammalian cells
  JOURNAL   Patent: US 5843776-A 2 01-DEC-1998;
FEATURES             Location/Qualifiers
     source          1. .13
                     /organism="unknown"
BASE COUNT        4 a      1 c      2 g      6 t
ORIGIN


  Query Match             100.0%;  Score 7;  DB 6;  Length 13;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+06;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TACTAAC 7
          |||||||
Db      7 TACTAAC 1


RESULT   9
I76088
LOCUS       I76088                  14 bp    DNA     linear   PAT 03-APR-1998
DEFINITION  Sequence 4 from patent US 5691137.
ACCESSION   I76088
VERSION     I76088.1  GI:3012242
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 14)
  AUTHORS   Rosbash,M. and Stutz,F.
  TITLE     Methods of screening candidate agents for biological activity using
            yeast cells
  JOURNAL   Patent: US 5691137-A 4 25-NOV-1997;
FEATURES             Location/Qualifiers
     source          1. .14
                     /organism="unknown"
BASE COUNT        5 a      5 c      0 g      4 t
ORIGIN


  Query Match             100.0%;  Score 7;  DB 6;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+06;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TACTAAC 7
          |||||||
Db      1 TACTAAC 7


RESULT  10
AR039499
LOCUS       AR039499                17 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 347 from patent US 5807743.
ACCESSION   AR039499
VERSION     AR039499.1  GI:5958862
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 17)
  AUTHORS   Stinchcomb,D.T. and McSwiggen,J.A.
  TITLE     Interleukin-2 receptor gamma-chain ribozymes
  JOURNAL   Patent: US 5807743-A 347 15-SEP-1998;
FEATURES             Location/Qualifiers
     source          1. .17
                     /organism="unknown"
BASE COUNT        5 a      2 c      3 g      7 t
ORIGIN


  Query Match             100.0%;  Score 7;  DB 6;  Length 17;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TACTAAC 7
          |||||||
Db      8 TACTAAC 14


RESULT  11
```

```
AR039501
LOCUS       AR039501                 17 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 349 from patent US 5807743.
ACCESSION   AR039501
VERSION     AR039501.1  GI:5958864
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 17)
  AUTHORS   Stinchcomb,D.T. and McSwiggen,J.A.
  TITLE     Interleukin-2 receptor gamma-chain ribozymes
  JOURNAL   Patent: US 5807743-A 349 15-SEP-1998;
FEATURES             Location/Qualifiers
     source          1. .17
                     /organism="unknown"
BASE COUNT        3 a      4 c      3 g      7 t
ORIGIN


  Query Match           100.0%;  Score 7;  DB 6;  Length 17;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TACTAAC 7
          |||||||
Db      5 TACTAAC 11


RESULT  12
AR039503
LOCUS       AR039503                 17 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 351 from patent US 5807743.
ACCESSION   AR039503
VERSION     AR039503.1  GI:5958866
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 17)
  AUTHORS   Stinchcomb,D.T. and McSwiggen,J.A.
  TITLE     Interleukin-2 receptor gamma-chain ribozymes
  JOURNAL   Patent: US 5807743-A 351 15-SEP-1998;
FEATURES             Location/Qualifiers
     source          1. .17
                     /organism="unknown"
BASE COUNT        4 a      5 c      0 g      8 t
ORIGIN


  Query Match           100.0%;  Score 7;  DB 6;  Length 17;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TACTAAC 7
          |||||||
Db      1 TACTAAC 7


RESULT  13
AR040467
LOCUS       AR040467                 17 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 1315 from patent US 5807743.
ACCESSION   AR040467
VERSION     AR040467.1  GI:5959830
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 17)
  AUTHORS   Stinchcomb,D.T. and McSwiggen,J.A.
  TITLE     Interleukin-2 receptor gamma-chain ribozymes
  JOURNAL   Patent: US 5807743-A 1315 15-SEP-1998;
FEATURES             Location/Qualifiers
     source          1. .17
                     /organism="unknown"
BASE COUNT        5 a      4 c      1 g      7 t
ORIGIN


  Query Match           100.0%;  Score 7;  DB 6;  Length 17;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TACTAAC 7
          |||||||
Db     11 TACTAAC 17


RESULT  14
AR040469
LOCUS       AR040469                 17 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 1317 from patent US 5807743.
ACCESSION   AR040469
VERSION     AR040469.1  GI:5959832
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 17)
  AUTHORS   Stinchcomb,D.T. and McSwiggen,J.A.
  TITLE     Interleukin-2 receptor gamma-chain ribozymes
  JOURNAL   Patent: US 5807743-A 1317 15-SEP-1998;
FEATURES             Location/Qualifiers
     source          1. .17
                     /organism="unknown"
BASE COUNT        5 a      3 c      1 g      8 t
ORIGIN


  Query Match           100.0%;  Score 7;  DB 6;  Length 17;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TACTAAC 7
          |||||||
Db      8 TACTAAC 14


RESULT  15
AR040471
LOCUS       AR040471                 17 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 1319 from patent US 5807743.
ACCESSION   AR040471
VERSION     AR040471.1  GI:5959834
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 17)
  AUTHORS   Stinchcomb,D.T. and McSwiggen,J.A.
  TITLE     Interleukin-2 receptor gamma-chain ribozymes
  JOURNAL   Patent: US 5807743-A 1319 15-SEP-1998;
FEATURES             Location/Qualifiers
     source          1. .17
                     /organism="unknown"
BASE COUNT        4 a      6 c      1 g      6 t
ORIGIN


  Query Match           100.0%;  Score 7;  DB 6;  Length 17;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TACTAAC 7
          |||||||
```

Db 5 TACTAAC 11

Search completed: July 21, 2002, 09:45:15
Job time: 12306 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2002, 09:55:18 ; Search time 467.25 Seconds
(without alignments)
25.722 Million cell updates/sec

Title: US-09-754-014-10_COPY_16_22
Perfect score: 7
Sequence: 1 TACTAAC 7

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	7	100.0	7	19	AAV64932	Yeast intron conse
	2	7	100.0	7	19	AAV43552	Insertion sequence
	3	7	100.0	7	19	AAV43560	Insertion sequence
c	4	7	100.0	10	21	AAZ78514	Human dendritic ce
c	5	7	100.0	10	21	AAZ81140	Metastatic breast
c	6	7	100.0	10	22	AAF41210	Yeast NORF gene SA
c	7	7	100.0	10	22	AAF41530	Yeast NORF gene SA
	8	7	100.0	10	22	AAF42338	Yeast NORF gene SA
c	9	7	100.0	10	22	AAF42417	Yeast NORF gene SA
	10	7	100.0	10	22	AAF43472	Yeast NORF gene SA
c	11	7	100.0	11	22	AAH55245	Genomic DNA methyl
	12	7	100.0	11	22	AAH55246	Genomic DNA methyl
c	13	7	100.0	11	22	AAH55253	Genomic DNA methyl
	14	7	100.0	11	22	AAH55254	Genomic DNA methyl
c	15	7	100.0	12	18	AAX28509	Target sequence fo
c	16	7	100.0	12	21	AAZ99828	Nucleotide sequenc
c	17	7	100.0	12	21	AAZ99829	Nucleotide sequenc
	18	7	100.0	12	23	ABH67674	Oligonucleotide pr
	19	7	100.0	12	23	ABH68017	Oligonucleotide pr
	20	7	100.0	12	23	ABH68161	Oligonucleotide pr
	21	7	100.0	12	23	ABH69292	Oligonucleotide pr
	22	7	100.0	12	23	ABH69440	Oligonucleotide pr
	23	7	100.0	12	23	ABH69448	Oligonucleotide pr
	24	7	100.0	12	23	ABH69976	Oligonucleotide pr
	25	7	100.0	12	23	ABH70202	Oligonucleotide pr
c	26	7	100.0	12	23	ABH71367	Oligonucleotide pr
	27	7	100.0	12	23	ABH71653	Oligonucleotide pr
	28	7	100.0	12	23	ABH72304	Oligonucleotide pr
	29	7	100.0	12	23	ABH72862	Oligonucleotide pr
	30	7	100.0	12	23	ABH73012	Oligonucleotide pr
	31	7	100.0	12	23	ABH74291	Oligonucleotide pr
	32	7	100.0	12	23	ABH74947	Oligonucleotide pr
c	33	7	100.0	12	23	ABH75567	Oligonucleotide pr
c	34	7	100.0	12	23	ABH76565	Oligonucleotide pr
c	35	7	100.0	12	23	ABH76876	Oligonucleotide pr
	36	7	100.0	12	23	ABH77257	Oligonucleotide pr
	37	7	100.0	12	23	ABH77374	Oligonucleotide pr
	38	7	100.0	12	23	ABH77375	Oligonucleotide pr
c	39	7	100.0	12	23	ABH77914	Oligonucleotide pr
	40	7	100.0	12	23	ABH78354	Oligonucleotide pr
	41	7	100.0	12	23	ABH78666	Oligonucleotide pr
	42	7	100.0	12	23	ABH78667	Oligonucleotide pr
	43	7	100.0	12	23	ABH78976	Oligonucleotide pr
	44	7	100.0	12	23	ABH79343	Oligonucleotide pr
	45	7	100.0	12	23	ABH79427	Oligonucleotide pr

ALIGNMENTS

RESULT 1
AAV64932
ID AAV64932 standard; RNA; 7 BP.
XX
AC AAV64932;
XX
DT 15-MAR-1999 (first entry)
XX
DE Yeast intron consensus branch site.
XX
KW Herpes simplex virus type-1; HSV-1; latency associated transcript;
KW LAT; LATin; gene transcript stabilisation; gene expression;
KW gene therapy; yeast; ss.
XX
OS Saccharomyces cerevisiae.
XX
FH Key Location/Qualifiers
FT misc_feature 6
FT /*tag= a
FT /note= "branchsite"
XX
PN WO9848004-A1.
XX
PD 29-OCT-1998.
XX
PF 17-APR-1998; 98WO-US07691.
XX
PR 18-APR-1997; 97US-0044664.
XX
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.
XX
PI Fraser NW, Krummenacher CF, Zabolotny JM;

```
XX
DR   WPI; 1998-609982/51.
XX
PT   Increasing expression of genes having unstable RNA transcripts,
PT   particularly for gene therapy - using a construct including gene
PT   flanked by intron fragments that include a hairpin next to the
PT   intron branchpoint
XX
PS   Example 7; Page 35; 106pp; English.
XX
CC   This is the nucleotide sequence of the yeast intron consensus
CC   branch site.  It was used as a control for determining which
CC   nucleotide within the branchpoint region of herpes simplex virus
CC   type 1 (HSV-1) latency associated transcript (LAT) is the nucleotide
CC   that forms a 2'-5' phosphodiester bond with the 5' splice donor
CC   site (see AAV64934).  The invention relates to methods of stabilising
CC   unstable gene transcripts.  A claimed polynucleotide comprises: (a)
CC   a polynucleotide encoding a gene product; (b) a 5'-sequence of an
CC   intron, including the splice donor and splice acceptor sites (see
CC   AAV64885-86); and (c) a 3'-sequence of the same intron, including a
CC   hairpin structure (see AAV64887) next to the intron's branchpoint.  A
CC   preferred intron is the 2.0 kb LAT of a herpes virus.  Methods and
CC   compositions using the polynucleotide can be used in gene therapy
CC   and more generally as research reagents, markers of gene production,
CC   in therapeutic or diagnostic compositions, in drug screening and to
CC   identify transcripts produced only at selected stages of the cell
CC   cycle.
XX
SQ   Sequence 7 BP; 3 A; 2 C; 0 G; 2 U; 0 other;

  Query Match             100.0%;  Score 7;  DB 19;  Length 7;
  Best Local Similarity   71.4%;  Pred. No. 2.4e+08;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TACTAAC 7
            :||:|||
Db       1 uacuaac 7


RESULT   2
AAV43552
ID   AAV43552 standard; DNA; 7 BP.
XX
AC   AAV43552;
XX
DT   16-SEP-1998  (first entry)
XX
DE   Insertion sequence 5 used for creating a tagged gene.
XX
KW   Tagged gene; tagged transcript; hybrid intron; protein tag;
KW   protein isolation; recombination; subcellular structure analysis;
KW   transcriptional regulation; viral infection; ss.
XX
OS   Synthetic.
OS   Unidentified.
XX
PN   WO9820031-A1.
XX
PD   14-MAY-1998.
XX
PF   07-NOV-1997;   97WO-US20150.
XX
PR   08-NOV-1996;   96US-0705404.
XX
PA   (JARV/) JARVIK J W.
XX
PI   Jarvik JW;
XX
DR   WPI; 1998-286861/25.
XX
PT   Tagging genes, transcripts and proteins - using tag-creating DNA
PT   inserted into intron of gene to create 2 hybrid introns separated by
PT   new exon encoding protein tag
XX
PS   Claim 1; Page 33; 66pp; English.
XX
CC   This sequence is used in the method of invention for tagging genes,
CC   transcripts and proteins in cells in a single recombinational event. The
CC   method comprises producing a tagged gene by inserting a DNA sequence
CC   into an intron of a gene by selecting a DNA sequence having a 5' portion
CC   free of any nucleotide sequence selected from AAV43548 to AAV43551, a
CC   nucleotide sequence selected from AAV43552 to AAV43560 and nucleotide
CC   sequences identical to a known splice branch site in a known gene,
CC   sequences identical in length to a known spacer region between splice
CC   branch and acceptor sites in a known gene, sequences identical to a known
CC   splice acceptor site in a known gene, sequence identical to a known
CC   splice donor site in a known gene, an open reading frame (ORF) 3N-1
CC   nucleotides in length, the ORF encoding a known peptide tag recognisable
CC   by a known reaction characteristic of the known peptide tag and sequences
CC   selected form CAGG and TAGG. The DNA sequence is inserted into the intron
CC   within the gene to create a tagged gene, and the tagged gene is incubated
CC   within a cell so as to maintain intact or to introduce the tagged gene
CC   within the genome of the cell. The method is used for isolating proteins,
CC   RNA and genes, for analysis of subcellular structures, cellular responses
CC   and transcriptional regulation, for the study of viral infection and for
CC   diagnosis of disease.
XX
SQ   Sequence 7 BP; 3 A; 2 C; 0 G; 2 T; 0 other;

  Query Match             100.0%;  Score 7;  DB 19;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+08;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TACTAAC 7
            |||||||
Db       1 tactaac 7


RESULT   3
AAV43560
ID   AAV43560 standard; DNA; 7 BP.
XX
AC   AAV43560;
XX
DT   16-SEP-1998  (first entry)
XX
DE   Insertion sequence 13 used for creating a tagged gene.
XX
KW   Tagged gene; tagged transcript; hybrid intron; protein tag;
KW   protein isolation; recombination; subcellular structure analysis;
KW   transcriptional regulation; viral infection; ss.
XX
OS   Synthetic.
OS   Unidentified.
XX
PN   WO9820031-A1.
XX
PD   14-MAY-1998.
XX
PF   07-NOV-1997;   97WO-US20150.
XX
PR   08-NOV-1996;   96US-0705404.
XX
PA   (JARV/) JARVIK J W.
XX
PI   Jarvik JW;
XX
DR   WPI; 1998-286861/25.
XX
PT   Tagging genes, transcripts and proteins - using tag-creating DNA
PT   inserted into intron of gene to create 2 hybrid introns separated by
PT   new exon encoding protein tag
XX
PS   Claim 1; Page 33; 66pp; English.
```

```
XX
CC   This sequence is used in the method of invention for tagging genes,
CC   transcripts and proteins in cells in a single recombinational event. The
CC   method comprises producing a tagged gene by inserting a DNA sequence
CC   into an intron of a gene by selecting a DNA sequence having a 5' portion
CC   free of any nucleotide sequence selected from AAV43548 to AAV43551, a
CC   nucleotide sequence selected from AAV43552 to AAV43560 and nucleotide
CC   sequences identical to a known splice branch site in a known gene,
CC   sequences identical in length to a known spacer region between splice
CC   branch and acceptor sites in a known gene, sequences identical to a known
CC   splice acceptor site in a known gene, sequence identical to a known
CC   splice donor site in a known gene, an open reading frame (ORF) 3N-1
CC   nucleotides in length, the ORF encoding a known peptide tag recognisable
CC   by a known reaction characteristic of the known peptide tag and sequences
CC   selected form CAGG and TAGG. The DNA sequence is inserted into the intron
CC   within the gene to create a tagged gene, and the tagged gene is incubated
CC   within a cell so as to maintain intact or to introduce the tagged gene
CC   within the genome of the cell. The method is used for isolating proteins,
CC   RNA and genes, for analysis of subcellular structures, cellular responses
CC   and transcriptional regulation, for the study of viral infection and for
CC   diagnosis of disease.
XX
SQ   Sequence 7 BP; 3 A; 2 C; 0 G; 2 T; 0 other;


  Query Match             100.0%;  Score 7;  DB 19;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+08;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TACTAAC 7
           |||||||
Db       1 tactaac 7


RESULT   4
AAZ78514/c
ID   AAZ78514 standard; DNA; 10 BP.
XX
AC   AAZ78514;
XX
DT   10-APR-2000  (first entry)
XX
DE   Human dendritic cell SAGE tag, SEQ ID NO:942.
XX
KW   SAGE tag; serial analysis of gene expression; antigen-presenting cell;
KW   APC; monocyte-derived dendritic cell; differential gene expression;
KW   immunostimulatory cofactor; costimulatory factor; CTL;
KW   cytotoxic T-lymphocyte; tumour antigen; immunotherapy; anticancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO9965924-A2.
XX
PD   23-DEC-1999.
XX
PF   18-JUN-1999;   99WO-US13800.
XX
PR   19-JUN-1998;   98US-0089833.
PR   19-JUN-1998;   98US-0089844.
PR   19-JUN-1998;   98US-0089853.
PR   19-JUN-1998;   98US-0089878.
PR   19-JUN-1998;   98US-0089991.
PR   19-JUN-1998;   98US-0089992.
PR   19-JUN-1998;   98US-0089993.
PR   19-JUN-1998;   98US-0089994.
PR   19-JUN-1998;   98US-0089997.
PR   19-JUN-1998;   98US-0089999.
PR   19-JUN-1998;   98US-0090000.
PR   19-JUN-1998;   98US-0090035.
PR   19-JUN-1998;   98US-0090036.
PR   19-JUN-1998;   98US-0090039.
PR   19-JUN-1998;   98US-0090040.
PR   19-JUN-1998;   98US-0090041.
PR   19-JUN-1998;   98US-0090042.
PR   19-JUN-1998;   98US-0090043.
PR   19-JUN-1998;   98US-0090044.
PR   19-JUN-1998;   98US-0090045.
PR   19-JUN-1998;   98US-0090047.
PR   19-JUN-1998;   98US-0090048.
PR   19-JUN-1998;   98US-0090072.
PR   19-JUN-1998;   98US-0090076.
PR   19-JUN-1998;   98US-0090077.
PR   19-JUN-1998;   98US-0090078.
PR   19-JUN-1998;   98US-0090079.
PR   19-JUN-1998;   98US-0090080.
PR   08-DEC-1998;   98US-0111715.
XX
PA   (GENZ ) GENZYME CORP.
PA   (ROBE/) ROBERTS B L.
PA   (SHAN/) SHANKARA S.
XX
PI   Roberts BL,  Shankara S;
XX
DR   WPI; 2000-106077/09.
XX
PT   Isolated polynucleotides differentially expressed in antigen-presenting
PT   cells, useful in gene vaccines against cancer -
XX
PS   Claim 1; Page 92; 130pp; English.
XX
CC   Sequences AAZ77573-Z79709 represent SAGE (serial analysis of gene
CC   expression) tags used to identify mRNA transcripts encoding
CC   immunostimulatory cofactor proteins which are preferentially or
CC   differentially expressed in monocyte-derived dendritic cells compared
CC   with monocytes. Some of the transcripts correspond to known genes or
CC   ESTs (expressed sequence tags) which were previously unknown to be
CC   preferentially or differentially expressed in dendritic cells, while
CC   other transcripts correspond to novel genes. Antigen-presenting cell
CC   (APC)-associated costimulatory factors play an important role in the
CC   activation of the cytotoxic immune response, particularly against tumour
CC   cells. Tumour antigen presentation via the MHC (major histocompatibility
CC   complex) and subsequent recognition by T-cell receptors is alone
CC   insufficient to activate a robust cytotoxic immune response that can
CC   lyse the tumour cells, immunostimulatory cofactors also being required
CC   for efficient activation of cytotoxic T-lymphocytes (CTLs). Nucleic acid
CC   sequences identified using the SAGE tags have several potential uses.
CC   They may be used in vaccines to induce an immune response, particularly
CC   against a tumour antigen; to modulate the genotype of an APC; to screen
CC   for agents that modulate expression of differentially expressed genes in
CC   an APC; and as hybridisation probes/amplification primers for the
CC   diagnosis, prognosis and monitoring of diseases related to abnormal
CC   expression of these genes. Detection of the dendritic cell
CC   differentially expressed genes, or of their encoded proteins, can be used
CC   to identify cells as belonging to the monocyte lineage. Cells containing
CC   these genes can be used in active immunotherapy (or to stimulate
CC   production of a population of antigen-specific effector cells) and
CC   vectors containing them are used in gene therapy. Co-administration of
CC   tumour antigens and APC-associated costimulatory factors ensures adequate
CC   antigen presentation to endogenous APCs and upregulates the APCs for the
CC   presentation of co-stimulatory signals, migration to T cell-rich sites,
CC   secretion of T cell growth factors and secretion of chemokines for
CC   recruitment of immune effector cells.
XX
SQ   Sequence 10 BP; 4 A; 1 C; 2 G; 3 T; 0 other;


  Query Match             100.0%;  Score 7;  DB 21;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 6.4e+04;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TACTAAC 7
           |||||||
Db       9 TACTAAC 3


RESULT   5
```

```
AAZ81140/c
ID   AAZ81140 standard; DNA; 10 BP.
XX
AC   AAZ81140;
XX
DT   07-APR-2000  (first entry)
XX
DE   Metastatic breast tumour cell upregulated transcript tag #374.
XX
KW   Human; metastatic breast tumour tissue; breast cancer; tag; primer;
KW   non-metastatic breast tumour tissue; gene therapy; anticancer;
KW   antimetastatic; vaccine; diagnosis; ss.
XX
OS   Homo sapiens.
XX
PN   WO9965928-A2.
XX
PD   23-DEC-1999.
XX
PF   18-JUN-1999;   99WO-US13647.
XX
PR   19-JUN-1998;   98US-0089853.
PR   19-JUN-1998;   98US-0089997.
PR   19-JUN-1998;   98US-0090039.
PR   19-JUN-1998;   98US-0090040.
PR   19-JUN-1998;   98US-0090041.
XX
PA   (GENZ ) GENZYME CORP.
PA   (ROBE/) ROBERTS B L.
PA   (SHAN/) SHANKARA S.
XX
PI   Roberts BL,  Shankara S;
XX
DR   WPI; 2000-106079/09.
XX
PT   Isolated polynucleotides differentially expressed between metastatic
PT   and non-metastatic breast cancer cells, useful for diagnosis,
PT   prevention and treatment of cancer -
XX
PS   Claim 1; Page 68; 219pp; English.
XX
CC   AAZ80767 to AAZ83941 represent tags corresponding to distinct
CC   transcripts that are preferentially transcribed in the metastatic breast
CC   tumour tissue (i.e. are upregulated in metastatic breast tumour cells).
CC   AAZ83942 to AAZ86677 represent tags corresponding to distinct transcripts
CC   that are preferentially transcribed in the primary or non-metastatic
CC   breast tumour tissue (i.e. are downregulated in metastatic breast tumour
CC   cells). These transcripts can be used for diagnosis, prognosis,
CC   monitoring and treatment of breast cancer, particularly where metastatic.
CC   Diagnosis is by standard immunoassays or hybridisation/amplification
CC   reactions. Compounds that modulate expression of the transcripts are
CC   potentially useful for treatment of (metastatic) breast cancer, while
CC   promoters from the transcripts are used to direct expression, in selected
CC   cell types, of e.g. therapeutic genes (also ribozymes or antisense
CC   sequences), particularly an antigen-encoding sequence for use in gene or
CC   cell-based vaccines. Polypeptides encoded by the transcripts are also
CC   useful in vaccines; for diagnosing breast cancer and for raising
CC   specific antibodies (Ab). Ab are used to detect the polypeptides or as
CC   therapeutic agents. Host cells that produce the polypeptides can be used
CC   to expand and isolate populations of educated, antigen-specific immune
CC   effecter cells, e.g. cytotoxic T lymphocytes, and these used for
CC   adoptive immunotherapy.
XX
SQ   Sequence 10 BP; 3 A; 0 C; 3 G; 4 T; 0 other;

  Query Match           100.0%;  Score 7;  DB 21;  Length 10;
  Best Local Similarity  100.0%;  Pred. No. 6.4e+04;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TACTAAC 7
            |||||||
Db      10 TACTAAC 4
```

```
RESULT   6
AAF41210/c
ID   AAF41210 standard; DNA; 10 BP.
XX
AC   AAF41210;
XX
DT   23-MAR-2001  (first entry)
XX
DE   Yeast NORF gene SAGE tag oligonucleotide SEQ ID NO:7949.
XX
KW   Yeast; Saccharomyces cerevisiae; characterisation; cell cycle; NORF;
KW   nor previously assigned open reading frame; nonannotated ORF; SAGE;
KW   serial analysis of gene expression; antifungal; tag; identification;
KW   linker; PCR primer; ds.
XX
OS   Saccharomyces cerevisiae.
XX
PN   WO200077214-A2.
XX
PD   21-DEC-2000.
XX
PF   14-JUN-2000; 2000WO-US16223.
XX
PR   16-JUN-1999;   99US-0335032.
XX
PA   (UYJO ) UNIV JOHNS HOPKINS.
XX
PI   Velculescu V,  Vogelstein B,  Kinzler K;
XX
DR   WPI; 2001-061874/07.
XX
PT   Yeast gene coding sequences comprising NORF genes with serial analysis
PT   of gene expression (SAGE) tags, useful for studying, monitoring and
PT   affecting phases of the cell cycle -
XX
PS   Example; Page 283; 419pp; English.
XX
CC   The present invention describes an isolated DNA molecule comprising a
CC   coding sequence of a yeast gene selected from a group of 745 NORF (not
CC   previously assigned open reading frame; or nonannotated ORF) genes
CC   comprising a SAGE (serial analysis of gene expression) tag. Also
CC   described are: (1) a method (M1) of using NORF genes to affect the cell
CC   cycle comprising administering a NORF gene whose expression varies by at
CC   least 10% between any two phases of the cell cycle selected from log
CC   phase, S phase and G2/M; (2) a method (M2) for screening candidate
CC   antifungal drugs comprising: (a) contacting a test substance with a
CC   yeast cell; and (b) monitoring expression of a NORF gene whose
CC   expression varies as in M1, where a test substance which modifies the
CC   expression of the yeast gene is a candidate antifungal drug; (3) a method
CC   (M3) for identifying human genes which are involved in cell cycle
CC   progression comprising contacting human DNA with a probe which comprises
CC   at least 10 contiguous nucleotides of a NORF gene whose expression varies
CC   as in M1; and (4) a method (M4) for identifying a candidate drug as a
CC   member of a class of drugs having a characteristic effect on gene
CC   expression in a yeast cell comprising contacting a yeast cell with a
CC   candidate drug and monitoring expression in the yeast cell of at least 1
CC   NORF gene whose expression is affected by the class of drugs. The NORF
CC   genes may be used to study, monitor and affect phases of the cell cycle,
CC   the differentially expressed genes may be used as markers of phases of
CC   the cell cycle. The methods may be used to identify candidate drugs which
CC   affect the cell cycle and for identification of antifungal drugs.
CC   AAF33268 to AAF44064 represent SAGE tags used in the exemplification of
CC   the present invention. AAF33262 to AAF33267 represent linkers and PCR
CC   primers used in the SAGE method, in the exemplification of the present
CC   invention.
XX
SQ   Sequence 10 BP; 2 A; 1 C; 3 G; 4 T; 0 other;

  Query Match           100.0%;  Score 7;  DB 22;  Length 10;
  Best Local Similarity  100.0%;  Pred. No. 6.4e+04;
```

```
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TACTAAC 7
           |||||||
Db       9 TACTAAC 3


RESULT   7
AAF41530/c
ID   AAF41530 standard; DNA; 10 BP.
XX
AC   AAF41530;
XX
DT   23-MAR-2001  (first entry)
XX
DE   Yeast NORF gene SAGE tag oligonucleotide SEQ ID NO:8269.
XX
KW   Yeast; Saccharomyces cerevisiae; characterisation; cell cycle; NORF;
KW   nor previously assigned open reading frame; nonannotated ORF; SAGE;
KW   serial analysis of gene expression; antifungal; tag; identification;
KW   linker; PCR primer; ds.
XX
OS   Saccharomyces cerevisiae.
XX
PN   WO200077214-A2.
XX
PD   21-DEC-2000.
XX
PF   14-JUN-2000; 2000WO-US16223.
XX
PR   16-JUN-1999;   99US-0335032.
XX
PA   (UYJO ) UNIV JOHNS HOPKINS.
XX
PI   Velculescu V,  Vogelstein B,  Kinzler K;
XX
DR   WPI; 2001-061874/07.
XX
PT   Yeast gene coding sequences comprising NORF genes with serial analysis
PT   of gene expression (SAGE) tags, useful for studying, monitoring and
PT   affecting phases of the cell cycle -
XX
PS   Example; Page 295; 419pp; English.
XX
CC   The present invention describes an isolated DNA molecule comprising a
CC   coding sequence of a yeast gene selected from a group of 745 NORF (not
CC   previously assigned open reading frame; or nonannotated ORF) genes
CC   comprising a SAGE (serial analysis of gene expression) tag. Also
CC   described are: (1) a method (M1) of using NORF genes to affect the cell
CC   cycle comprising administering a NORF gene whose expression varies by at
CC   least 10% between any two phases of the cell cycle selected from log
CC   phase, S phase and G2/M; (2) a method (M2) for screening candidate
CC   antifungal drugs comprising: (a) contacting a test substance with a
CC   yeast cell; and (b) monitoring expression of a NORF gene whose
CC   expression varies as in M1, where a test substance which modifies the
CC   expression of the yeast gene is a candidate antifungal drug; (3) a method
CC   (M3) for identifying human genes which are involved in cell cycle
CC   progression comprising contacting human DNA with a probe which comprises
CC   at least 10 contiguous nucleotides of a NORF gene whose expression varies
CC   as in M1; and (4) a method (M4) for identifying a candidate drug as a
CC   member of a class of drugs having a characteristic effect on gene
CC   expression in a yeast cell comprising contacting a yeast cell with a
CC   candidate drug and monitoring expression in the yeast cell of at least 1
CC   NORF gene whose expression is affected by the class of drugs. The NORF
CC   genes may be used to study, monitor and affect phases of the cell cycle,
CC   the differentially expressed genes may be used as markers of phases of
CC   the cell cycle. The methods may be used to identify candidate drugs which
CC   affect the cell cycle and for identification of antifungal drugs.
CC   AAF33268 to AAF44064 represent SAGE tags used in the exemplification of
CC   the present invention. AAF33262 to AAF33267 represent linkers and PCR
CC   primers used in the SAGE method, in the exemplification of the present
CC   invention.
XX
SQ   Sequence 10 BP; 2 A; 1 C; 2 G; 5 T; 0 other;


  Query Match             100.0%;  Score 7;  DB 22;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 6.4e+04;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TACTAAC 7
           |||||||
Db       9 TACTAAC 3


RESULT   8
AAF42338
ID   AAF42338 standard; DNA; 10 BP.
XX
AC   AAF42338;
XX
DT   23-MAR-2001  (first entry)
XX
DE   Yeast NORF gene SAGE tag oligonucleotide SEQ ID NO:9077.
XX
KW   Yeast; Saccharomyces cerevisiae; characterisation; cell cycle; NORF;
KW   nor previously assigned open reading frame; nonannotated ORF; SAGE;
KW   serial analysis of gene expression; antifungal; tag; identification;
KW   linker; PCR primer; ds.
XX
OS   Saccharomyces cerevisiae.
XX
PN   WO200077214-A2.
XX
PD   21-DEC-2000.
XX
PF   14-JUN-2000; 2000WO-US16223.
XX
PR   16-JUN-1999;   99US-0335032.
XX
PA   (UYJO ) UNIV JOHNS HOPKINS.
XX
PI   Velculescu V,  Vogelstein B,  Kinzler K;
XX
DR   WPI; 2001-061874/07.
XX
PT   Yeast gene coding sequences comprising NORF genes with serial analysis
PT   of gene expression (SAGE) tags, useful for studying, monitoring and
PT   affecting phases of the cell cycle -
XX
PS   Example; Page 324; 419pp; English.
XX
CC   The present invention describes an isolated DNA molecule comprising a
CC   coding sequence of a yeast gene selected from a group of 745 NORF (not
CC   previously assigned open reading frame; or nonannotated ORF) genes
CC   comprising a SAGE (serial analysis of gene expression) tag. Also
CC   described are: (1) a method (M1) of using NORF genes to affect the cell
CC   cycle comprising administering a NORF gene whose expression varies by at
CC   least 10% between any two phases of the cell cycle selected from log
CC   phase, S phase and G2/M; (2) a method (M2) for screening candidate
CC   antifungal drugs comprising: (a) contacting a test substance with a
CC   yeast cell; and (b) monitoring expression of a NORF gene whose
CC   expression varies as in M1, where a test substance which modifies the
CC   expression of the yeast gene is a candidate antifungal drug; (3) a method
CC   (M3) for identifying human genes which are involved in cell cycle
CC   progression comprising contacting human DNA with a probe which comprises
CC   at least 10 contiguous nucleotides of a NORF gene whose expression varies
CC   as in M1; and (4) a method (M4) for identifying a candidate drug as a
CC   member of a class of drugs having a characteristic effect on gene
CC   expression in a yeast cell comprising contacting a yeast cell with a
CC   candidate drug and monitoring expression in the yeast cell of at least 1
CC   NORF gene whose expression is affected by the class of drugs. The NORF
CC   genes may be used to study, monitor and affect phases of the cell cycle,
CC   the differentially expressed genes may be used as markers of phases of
CC   the cell cycle. The methods may be used to identify candidate drugs which
CC   affect the cell cycle and for identification of antifungal drugs.
```

```
CC   AAF33268 to AAF44064 represent SAGE tags used in the exemplification of
CC   the present invention. AAF33262 to AAF33267 represent linkers and PCR
CC   primers used in the SAGE method, in the exemplification of the present
CC   invention.
XX
SQ   Sequence 10 BP; 4 A; 3 C; 1 G; 2 T; 0 other;


  Query Match             100.0%;  Score 7;  DB 22;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 6.4e+04;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TACTAAC 7
           |||||||
Db       4 tactaac 10


RESULT   9
AAF42417/c
ID   AAF42417 standard; DNA; 10 BP.
XX
AC   AAF42417;
XX
DT   23-MAR-2001  (first entry)
XX
DE   Yeast NORF gene SAGE tag oligonucleotide SEQ ID NO:9156.
XX
KW   Yeast; Saccharomyces cerevisiae; characterisation; cell cycle; NORF;
KW   nor previously assigned open reading frame; nonannotated ORF; SAGE;
KW   serial analysis of gene expression; antifungal; tag; identification;
KW   linker; PCR primer; ds.
XX
OS   Saccharomyces cerevisiae.
XX
PN   WO200077214-A2.
XX
PD   21-DEC-2000.
XX
PF   14-JUN-2000; 2000WO-US16223.
XX
PR   16-JUN-1999;   99US-0335032.
XX
PA   (UYJO ) UNIV JOHNS HOPKINS.
XX
PI   Velculescu V,  Vogelstein B,  Kinzler K;
XX
DR   WPI; 2001-061874/07.
XX
PT   Yeast gene coding sequences comprising NORF genes with serial analysis
PT   of gene expression (SAGE) tags, useful for studying, monitoring and
PT   affecting phases of the cell cycle -
XX
PS   Example; Page 327; 419pp; English.
XX
CC   The present invention describes an isolated DNA molecule comprising a
CC   coding sequence of a yeast gene selected from a group of 745 NORF (not
CC   previously assigned open reading frame; or nonannotated ORF) genes
CC   comprising a SAGE (serial analysis of gene expression) tag. Also
CC   described are: (1) a method (M1) of using NORF genes to affect the cell
CC   cycle comprising administering a NORF gene whose expression varies by at
CC   least 10% between any two phases of the cell cycle selected from log
CC   phase, S phase and G2/M; (2) a method (M2) for screening candidate
CC   antifungal drugs comprising: (a) contacting a test substance with a
CC   yeast cell; and (b) monitoring expression of a NORF gene whose
CC   expression varies as in M1, where a test substance which modifies the
CC   expression of the yeast gene is a candidate antifungal drug; (3) a method
CC   (M3) for identifying human genes which are involved in cell cycle
CC   progression comprising contacting human DNA with a probe which comprises
CC   at least 10 contiguous nucleotides of a NORF gene whose expression varies
CC   as in M1; and (4) a method (M4) for identifying a candidate drug as a
CC   member of a class of drugs having a characteristic effect on gene
CC   expression in a yeast cell comprising contacting a yeast cell with a
CC   candidate drug and monitoring expression in the yeast cell of at least 1
CC   NORF gene whose expression is affected by the class of drugs. The NORF
CC   genes may be used to study, monitor and affect phases of the cell cycle,
CC   the differentially expressed genes may be used as markers of phases of
CC   the cell cycle. The methods may be used to identify candidate drugs which
CC   affect the cell cycle and for identification of antifungal drugs.
CC   AAF33268 to AAF44064 represent SAGE tags used in the exemplification of
CC   the present invention. AAF33262 to AAF33267 represent linkers and PCR
CC   primers used in the SAGE method, in the exemplification of the present
CC   invention.
XX
SQ   Sequence 10 BP; 2 A; 2 C; 3 G; 3 T; 0 other;


  Query Match             100.0%;  Score 7;  DB 22;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 6.4e+04;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TACTAAC 7
           |||||||
Db      10 TACTAAC 4


RESULT  10
AAF43472
ID   AAF43472 standard; DNA; 10 BP.
XX
AC   AAF43472;
XX
DT   23-MAR-2001  (first entry)
XX
DE   Yeast NORF gene SAGE tag oligonucleotide SEQ ID NO:11611.
XX
KW   Yeast; Saccharomyces cerevisiae; characterisation; cell cycle; NORF;
KW   nor previously assigned open reading frame; nonannotated ORF; SAGE;
KW   serial analysis of gene expression; antifungal; tag; identification;
KW   linker; PCR primer; ds.
XX
OS   Saccharomyces cerevisiae.
XX
PN   WO200077214-A2.
XX
PD   21-DEC-2000.
XX
PF   14-JUN-2000; 2000WO-US16223.
XX
PR   16-JUN-1999;   99US-0335032.
XX
PA   (UYJO ) UNIV JOHNS HOPKINS.
XX
PI   Velculescu V,  Vogelstein B,  Kinzler K;
XX
DR   WPI; 2001-061874/07.
XX
PT   Yeast gene coding sequences comprising NORF genes with serial analysis
PT   of gene expression (SAGE) tags, useful for studying, monitoring and
PT   affecting phases of the cell cycle -
XX
PS   Example; Page 364; 419pp; English.
XX
CC   The present invention describes an isolated DNA molecule comprising a
CC   coding sequence of a yeast gene selected from a group of 745 NORF (not
CC   previously assigned open reading frame; or nonannotated ORF) genes
CC   comprising a SAGE (serial analysis of gene expression) tag. Also
CC   described are: (1) a method (M1) of using NORF genes to affect the cell
CC   cycle comprising administering a NORF gene whose expression varies by at
CC   least 10% between any two phases of the cell cycle selected from log
CC   phase, S phase and G2/M; (2) a method (M2) for screening candidate
CC   antifungal drugs comprising: (a) contacting a test substance with a
CC   yeast cell; and (b) monitoring expression of a NORF gene whose
CC   expression varies as in M1, where a test substance which modifies the
CC   expression of the yeast gene is a candidate antifungal drug; (3) a method
CC   (M3) for identifying human genes which are involved in cell cycle
CC   progression comprising contacting human DNA with a probe which comprises
```

CC at least 10 contiguous nucleotides of a NORF gene whose expression varies
CC as in M1; and (4) a method (M4) for identifying a candidate drug as a
CC member of a class of drugs having a characteristic effect on gene
CC expression in a yeast cell comprising contacting a yeast cell with a
CC candidate drug and monitoring expression in the yeast cell of at least 1
CC NORF gene whose expression is affected by the class of drugs. The NORF
CC genes may be used to study, monitor and affect phases of the cell cycle,
CC the differentially expressed genes may be used as markers of phases of
CC the cell cycle. The methods may be used to identify candidate drugs which
CC affect the cell cycle and for identification of antifungal drugs.
CC AAF33268 to AAF44064 represent SAGE tags used in the exemplification of
CC the present invention. AAF33262 to AAF33267 represent linkers and PCR
CC primers used in the SAGE method, in the exemplification of the present
CC invention.
XX
SQ Sequence 10 BP; 4 A; 2 C; 0 G; 4 T; 0 other;

Query Match 100.0%; Score 7; DB 22; Length 10;
Best Local Similarity 100.0%; Pred. No. 6.4e+04;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TACTAAC 7
|||||||
Db 2 tactaac 8

RESULT 11
AAH55245/c
ID AAH55245 standard; DNA; 11 BP.
XX
AC AAH55245;
XX
DT 03-SEP-2001 (first entry)
XX
DE Genomic DNA methylation parallel detection associated DNA fragment #147.
XX
KW DNA methylation; parallel detection; 5-unmethylated cytosine; CpG;
KW CpNpG; amplification; transcription regulation; genetic imprinting;
KW tumorigenesis; primer; ss.
XX
OS Unidentified.
XX
PN WO200142493-A2.
XX
PD 14-JUN-2001.
XX
PF 06-DEC-2000; 2000WO-DE04381.
XX
PR 06-DEC-1999; 99DE-1059691.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C;
XX
DR WPI; 2001-381705/40.
XX
PT Parallel detection of the methylation pattern of many genomic DNA
PT regions, useful for detecting aberrant methylation, includes multiple
PT amplification of chemically modified DNA -
XX
PS Claim 18; Page 21; 63pp; German.
XX
CC This invention describes a novel method for the parallel detection of the
CC methylation status of genomic DNA (I) which involves a (I) sample being
CC treated chemically to convert 5-unmethylated cytosine to uracil,
CC thymidine or some other base having hybridization behavior different from
CC that of C, then amplifying simultaneously at least 10 different fragments
CC (of fewer than 2 kb) using synthetic oligonucleotide (ON) primers. These
CC primers are based on regulatory, transcribed and/or translated segments
CC present in the sample after chemical treatment. The sequence context of
CC all, or some, of the CpG and CpNpG motifs in the amplified products is
CC then determined. The method is used to detect aberrant methylation
CC patterns in the genome, these are implicated in regulation of
CC transcription, genetic imprinting and tumorigenesis. Many target regions
CC in the genome can be analyzed simultaneously and it is not essential to
CC know the sequence context of all targeted regions. Primers may be
CC designed for preferential amplification of particular segments of
CC interest (e.g. promoters and exons).
XX
SQ Sequence 11 BP; 3 A; 0 C; 4 G; 4 T; 0 other;

Query Match 100.0%; Score 7; DB 22; Length 11;
Best Local Similarity 100.0%; Pred. No. 6.3e+04;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TACTAAC 7
|||||||
Db 11 TACTAAC 5

RESULT 12
AAH55246
ID AAH55246 standard; DNA; 11 BP.
XX
AC AAH55246;
XX
DT 03-SEP-2001 (first entry)
XX
DE Genomic DNA methylation parallel detection associated DNA fragment #148.
XX
KW DNA methylation; parallel detection; 5-unmethylated cytosine; CpG;
KW CpNpG; amplification; transcription regulation; genetic imprinting;
KW tumorigenesis; primer; ss.
XX
OS Unidentified.
XX
PN WO200142493-A2.
XX
PD 14-JUN-2001.
XX
PF 06-DEC-2000; 2000WO-DE04381.
XX
PR 06-DEC-1999; 99DE-1059691.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C;
XX
DR WPI; 2001-381705/40.
XX
PT Parallel detection of the methylation pattern of many genomic DNA
PT regions, useful for detecting aberrant methylation, includes multiple
PT amplification of chemically modified DNA -
XX
PS Claim 18; Page 21; 63pp; German.
XX
CC This invention describes a novel method for the parallel detection of the
CC methylation status of genomic DNA (I) which involves a (I) sample being
CC treated chemically to convert 5-unmethylated cytosine to uracil,
CC thymidine or some other base having hybridization behavior different from
CC that of C, then amplifying simultaneously at least 10 different fragments
CC (of fewer than 2 kb) using synthetic oligonucleotide (ON) primers. These
CC primers are based on regulatory, transcribed and/or translated segments
CC present in the sample after chemical treatment. The sequence context of
CC all, or some, of the CpG and CpNpG motifs in the amplified products is
CC then determined. The method is used to detect aberrant methylation
CC patterns in the genome, these are implicated in regulation of
CC transcription, genetic imprinting and tumorigenesis. Many target regions
CC in the genome can be analyzed simultaneously and it is not essential to
CC know the sequence context of all targeted regions. Primers may be
CC designed for preferential amplification of particular segments of
CC interest (e.g. promoters and exons).
XX
SQ Sequence 11 BP; 4 A; 4 C; 0 G; 3 T; 0 other;

```
  Query Match            100.0%;  Score 7;  DB 22;  Length 11;
  Best Local Similarity  100.0%;  Pred. No. 6.3e+04;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TACTAAC 7
          |||||||
Db      1 tactaac 7


RESULT  13
AAH55253/c
ID   AAH55253 standard; DNA; 11 BP.
XX
AC   AAH55253;
XX
DT   03-SEP-2001  (first entry)
XX
DE   Genomic DNA methylation parallel detection associated DNA fragment #155.
XX
KW   DNA methylation; parallel detection; 5-unmethylated cytosine; CpG;
KW   CpNpG; amplification; transcription regulation; genetic imprinting;
KW   tumorigenesis; primer; ss.
XX
OS   Unidentified.
XX
PN   WO200142493-A2.
XX
PD   14-JUN-2001.
XX
PF   06-DEC-2000; 2000WO-DE04381.
XX
PR   06-DEC-1999;   99DE-1059691.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C;
XX
DR   WPI; 2001-381705/40.
XX
PT   Parallel detection of the methylation pattern of many genomic DNA
PT   regions, useful for detecting aberrant methylation, includes multiple
PT   amplification of chemically modified DNA  -
XX
PS   Claim 18; Page 21; 63pp; German.
XX
CC   This invention describes a novel method for the parallel detection of the
CC   methylation status of genomic DNA (I) which involves a (I) sample being
CC   treated chemically to convert 5-unmethylated cytosine to uracil,
CC   thymidine or some other base having hybridization behavior different from
CC   that of C, then amplifying simultaneously at least 10 different fragments
CC   (of fewer than 2 kb) using synthetic oligonucleotide (ON) primers. These
CC   primers are based on regulatory, transcribed and/or translated segments
CC   present in the sample after chemical treatment. The sequence context of
CC   all, or some, of the CpG and CpNpG motifs in the amplified products is
CC   then determined. The method is used to detect aberrant methylation
CC   patterns in the genome, these are implicated in regulation of
CC   transcription, genetic imprinting and tumorigenesis. Many target regions
CC   in the genome can be analyzed simultaneously and it is not essential to
CC   know the sequence context of all targeted regions. Primers may be
CC   designed for preferential amplification of particular segments of
CC   interest (e.g. promoters and exons).
XX
SQ   Sequence 11 BP; 3 A; 0 C; 4 G; 4 T; 0 other;


  Query Match            100.0%;  Score 7;  DB 22;  Length 11;
  Best Local Similarity  100.0%;  Pred. No. 6.3e+04;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TACTAAC 7
          |||||||
Db      11 TACTAAC 5


RESULT  14
AAH55254
ID   AAH55254 standard; DNA; 11 BP.
XX
AC   AAH55254;
XX
DT   03-SEP-2001  (first entry)
XX
DE   Genomic DNA methylation parallel detection associated DNA fragment #156.
XX
KW   DNA methylation; parallel detection; 5-unmethylated cytosine; CpG;
KW   CpNpG; amplification; transcription regulation; genetic imprinting;
KW   tumorigenesis; primer; ss.
XX
OS   Unidentified.
XX
PN   WO200142493-A2.
XX
PD   14-JUN-2001.
XX
PF   06-DEC-2000; 2000WO-DE04381.
XX
PR   06-DEC-1999;   99DE-1059691.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C;
XX
DR   WPI; 2001-381705/40.
XX
PT   Parallel detection of the methylation pattern of many genomic DNA
PT   regions, useful for detecting aberrant methylation, includes multiple
PT   amplification of chemically modified DNA  -
XX
PS   Claim 18; Page 21; 63pp; German.
XX
CC   This invention describes a novel method for the parallel detection of the
CC   methylation status of genomic DNA (I) which involves a (I) sample being
CC   treated chemically to convert 5-unmethylated cytosine to uracil,
CC   thymidine or some other base having hybridization behavior different from
CC   that of C, then amplifying simultaneously at least 10 different fragments
CC   (of fewer than 2 kb) using synthetic oligonucleotide (ON) primers. These
CC   primers are based on regulatory, transcribed and/or translated segments
CC   present in the sample after chemical treatment. The sequence context of
CC   all, or some, of the CpG and CpNpG motifs in the amplified products is
CC   then determined. The method is used to detect aberrant methylation
CC   patterns in the genome, these are implicated in regulation of
CC   transcription, genetic imprinting and tumorigenesis. Many target regions
CC   in the genome can be analyzed simultaneously and it is not essential to
CC   know the sequence context of all targeted regions. Primers may be
CC   designed for preferential amplification of particular segments of
CC   interest (e.g. promoters and exons).
XX
SQ   Sequence 11 BP; 4 A; 4 C; 0 G; 3 T; 0 other;


  Query Match            100.0%;  Score 7;  DB 22;  Length 11;
  Best Local Similarity  100.0%;  Pred. No. 6.3e+04;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TACTAAC 7
          |||||||
Db      1 tactaac 7


RESULT  15
AAX28509/c
ID   AAX28509 standard; DNA; 12 BP.
XX
AC   AAX28509;
```

XX
DT 08-JUN-1999 (first entry)
XX
DE Target sequence for minor groove binding polyamide.
XX
KW Netropsin; Distamycin A analogue; polypyrrole; polyimidazole;
KW carboxamide; polyamide; minor groove binding; oligonucleotide;
KW conjugate; ds.
XX
OS Synthetic.
XX
PN WO9730975-A2.
XX
PD 28-AUG-1997.
XX
PF 20-FEB-1997; 97WO-US03332.
XX
PR 26-FEB-1996; 96US-0607078.
XX
PA (CALY) CALIFORNIA INST OF TECHNOLOGY.
XX
PI Baird EE, Dervan PB;
XX
DR WPI; 1997-435067/40.
XX
PT Preparation of poly-pyrrole and poly-imidazole carboxamides - and
PT production of polyamide-protein and polyamide-oligo:nucleotide
PT conjugates on solid supports
XX
PS Disclosure; Page 14; 167pp; English.
XX
CC The patent describes methods for the preparation of polyamides
CC containing imidazole and pyrrole carboxamides, and also their
CC conjugates with oligonucleotides and proteins. The processes
CC may be used e.g. for solid phase synthesis of analogues of the
CC di- and tri-N-methylpyrrole carboxamide antiviral antibiotics
CC Netropsin and Distamycin A. Materials may be produced which
CC recognise double stranded DNA by interaction with the minor
CC groove of the DNA. These materials may be used as antiviral,
CC antibacterial and antitumour agents. They may be used in design
CC of therapeutic agents. They may be used to bind/cleave double
CC stranded DNA at specific sites using iron and EDTA. The methods
CC give the polyamides and conjugates with high stepwise coupling
CC yields and give highly pure products.
XX
SQ Sequence 12 BP; 2 A; 2 C; 3 G; 5 T; 0 other;

Query Match 100.0%; Score 7; DB 18; Length 12;
Best Local Similarity 100.0%; Pred. No. 6.3e+04;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 TACTAAC 7
            |||||||
Db      9 TACTAAC 3
```

Search completed: July 21, 2002, 09:55:19
Job time: 6380 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2002, 09:47:18 ; Search time 112.48 Seconds
(without alignments)
15.287 Million cell updates/sec

Title: US-09-754-014-10_COPY_16_22
Perfect score: 7
Sequence: 1 TACTAAC 7

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	7	100.0	7	3	US-09-403-267-1	Sequence 1, Appli
	2	7	100.0	9	3	US-08-646-789A-63	Sequence 63, Appl
c	3	7	100.0	12	3	US-08-607-078-5	Sequence 5, Appli
	4	7	100.0	12	3	US-09-403-267-33	Sequence 33, Appl
c	5	7	100.0	13	1	US-08-148-058A-2	Sequence 2, Appli
c	6	7	100.0	13	1	US-08-478-042-2	Sequence 2, Appli
c	7	7	100.0	13	1	US-08-645-215-2	Sequence 2, Appli
c	8	7	100.0	13	2	US-08-466-604-2	Sequence 2, Appli
	9	7	100.0	14	1	US-08-297-808A-4	Sequence 4, Appli
	10	7	100.0	15	4	US-09-242-690A-16	Sequence 16, Appl
	11	7	100.0	17	1	US-07-990-965-3	Sequence 3, Appli
	12	7	100.0	17	1	US-08-758-306-347	Sequence 347, App
	13	7	100.0	17	1	US-08-758-306-349	Sequence 349, App
	14	7	100.0	17	1	US-08-758-306-351	Sequence 351, App
	15	7	100.0	17	1	US-08-758-306-1315	Sequence 1315, Ap
	16	7	100.0	17	1	US-08-758-306-1317	Sequence 1317, Ap
	17	7	100.0	17	1	US-08-758-306-1319	Sequence 1319, Ap
	18	7	100.0	17	1	US-08-758-306-1321	Sequence 1321, Ap
c	19	7	100.0	18	2	US-08-683-743-19	Sequence 19, Appl
	20	7	100.0	18	2	US-08-810-599-65	Sequence 65, Appl
c	21	7	100.0	18	3	US-08-784-582-63	Sequence 63, Appl
	22	7	100.0	18	4	US-08-413-740A-152	Sequence 152, App
	23	7	100.0	18	5	PCT-US95-04063-152	Sequence 152, App
	24	7	100.0	19	1	US-08-219-842-24	Sequence 24, Appl
	25	7	100.0	19	1	US-08-066-325-20	Sequence 20, Appl
	26	7	100.0	19	1	US-08-451-096-24	Sequence 24, Appl
c	27	7	100.0	19	2	US-08-483-695-15	Sequence 15, Appl
c	28	7	100.0	19	2	US-07-965-285-15	Sequence 15, Appl
c	29	7	100.0	19	2	US-08-487-231-15	Sequence 15, Appl
c	30	7	100.0	19	4	US-09-201-912-15	Sequence 15, Appl
c	31	7	100.0	19	4	US-09-338-907-477	Sequence 477, App
c	32	7	100.0	19	4	US-09-218-207-477	Sequence 477, App
c	33	7	100.0	20	1	US-08-229-145-15	Sequence 15, Appl
c	34	7	100.0	20	1	US-08-229-145-16	Sequence 16, Appl
c	35	7	100.0	20	1	US-08-466-285-7	Sequence 7, Appli
	36	7	100.0	20	1	US-08-647-584-123	Sequence 123, App
	37	7	100.0	20	2	US-08-639-501-28	Sequence 28, Appl
	38	7	100.0	20	3	US-09-044-946-28	Sequence 28, Appl
	39	7	100.0	20	3	US-09-166-186-184	Sequence 184, App
	40	7	100.0	20	3	US-09-166-186-185	Sequence 185, App
	41	7	100.0	20	3	US-09-044-908-28	Sequence 28, Appl
c	42	7	100.0	20	3	US-09-288-461-76	Sequence 76, Appl
	43	7	100.0	20	4	US-09-313-932-184	Sequence 184, App
	44	7	100.0	20	4	US-09-313-932-185	Sequence 185, App
	45	7	100.0	20	4	US-08-397-220B-34	Sequence 34, Appl

ALIGNMENTS

```
RESULT   1
US-09-403-267-1
; Sequence 1, Application US/09403267
; Patent No. 6159710
;  GENERAL INFORMATION:
;    APPLICANT:  Wistar Institute of Anatomy, and Biology
;    APPLICANT:  Fraser, Nigel W.
;    APPLICANT:  Zabolotny, Janice M.
;    APPLICANT:  Krummenacher, Claude F.
;    TITLE OF INVENTION:  Method and Compositions for Stabilizing
;    TITLE OF INVENTION:  Unstable Gene Transcripts
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howson and Howson
;      STREET:  Spring House Corporate Cntr., P.O. Box 457
;      CITY:  Spring House
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19477
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/403,267
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/044,664
;      FILING DATE:  18-APR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bak, Mary E.
;      REGISTRATION NUMBER:  31,215
;      REFERENCE/DOCKET NUMBER:  WST78APCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-540-9200
;      TELEFAX:  215-540-5818
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  RNA (genomic)
US-09-403-267-1


  Query Match             100.0%;  Score 7;  DB 3;  Length 7;
  Best Local Similarity   71.4%;  Pred. No. 3.3e+07;
```

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 TACTAAC 7
            :||:|||
Db      1 UACUAAC 7
```

```
RESULT   2
US-08-646-789A-63
; Sequence 63, Application US/08646789A
; Patent No. 6022863
;  GENERAL INFORMATION:
;    APPLICANT:  Peyman, John A.
;    TITLE OF INVENTION:  REGULATION OF GENE EXPRESSION
;    NUMBER OF SEQUENCES:  101
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/646,789A
;      FILING DATE:  May 21, 1996
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  6523-006
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-9741/8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  63:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-08-646-789A-63
```

Query Match 100.0%; Score 7; DB 3; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.5e+07;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 TACTAAC 7
            |||||||
Db      2 TACTAAC 8
```

```
RESULT   3
US-08-607-078-5/c
; Sequence 5, Application US/08607078
; Patent No. 6090947
;  GENERAL INFORMATION:
;    APPLICANT:  California Institute of Technology
;    TITLE OF INVENTION:  Method for the Synthesis of Pyrrole
;    TITLE OF INVENTION:  and Imidazole Carboxamides on a
;    TITLE OF INVENTION:  Solid Support
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Swanson & Bratschun, L.L.C.
;      STREET:  8400 E. Prentice Avenue, Suite 200
;      CITY:  Englewood
;      STATE:  Colorado
;      COUNTRY:  USA
;      ZIP:  80111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3 1/2 diskette, 1.44 MG
;      COMPUTER:  IBM pc compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WordPerfect 6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/607,078
;      FILING DATE:  February 26, 1996
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rosemary P. Kellogg
;      REGISTRATION NUMBER:  39,726
;      REFERENCE/DOCKET NUMBER:  CIT 2347
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 793-3333
;      TELEFAX:  (303) 793-3433
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-607-078-5
```

Query Match 100.0%; Score 7; DB 3; Length 12;
Best Local Similarity 100.0%; Pred. No. 6.7e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 TACTAAC 7
            |||||||
Db      9 TACTAAC 3
```

```
RESULT   4
US-09-403-267-33
; Sequence 33, Application US/09403267
; Patent No. 6159710
;  GENERAL INFORMATION:
;    APPLICANT:  Wistar Institute of Anatomy, and Biology
;    APPLICANT:  Fraser, Nigel W.
;    APPLICANT:  Zabolotny, Janice M.
;    APPLICANT:  Krummenacher, Claude F.
;    TITLE OF INVENTION:  Method and Compositions for Stabilizing
;    TITLE OF INVENTION:  Unstable Gene Transcripts
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howson and Howson
;      STREET:  Spring House Corporate Cntr., P.O. Box 457
;      CITY:  Spring House
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19477
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/403,267
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/044,664
;      FILING DATE:  18-APR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bak, Mary E.
;      REGISTRATION NUMBER:  31,215
```

```
;      REFERENCE/DOCKET NUMBER:  WST78APCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-540-9200
;      TELEFAX:  215-540-5818
;  INFORMATION FOR SEQ ID NO:  33:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-09-403-267-33


  Query Match           100.0%;  Score 7;  DB 3;  Length 12;
  Best Local Similarity  100.0%;  Pred. No. 6.7e+03;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TACTAAC 7
          |||||||
Db      6 TACTAAC 12


RESULT   5
US-08-148-058A-2/c
; Sequence 2, Application US/08148058A
; Patent No. 5804407
;  GENERAL INFORMATION:
;    APPLICANT:  TAMAOKI, TAIKI
;    APPLICANT:  NAKABAYASHI, HIDEKAZU
;    TITLE OF INVENTION:  IMPROVED METHOD OF EXPRESSING GENES IN
;    TITLE OF INVENTION:  MAMMALIAN CELLS
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BURNS, DOANE, SWECKER & MATHIS
;      STREET:  699 PRINCE STREET
;      CITY:  ALEXANDRIA
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/148,058A
;      FILING DATE:  04-NOV-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MOOI, LESLIE A.
;      REGISTRATION NUMBER:  37,047
;      REFERENCE/DOCKET NUMBER:  028722-074
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-854-7400
;      TELEFAX:  415-854-8275
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  13 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-148-058A-2


  Query Match           100.0%;  Score 7;  DB 1;  Length 13;
  Best Local Similarity  100.0%;  Pred. No. 6.6e+03;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TACTAAC 7
          |||||||
Db      7 TACTAAC 1


RESULT   6
US-08-478-042-2/c
; Sequence 2, Application US/08478042
; Patent No. 5807738
;  GENERAL INFORMATION:
;    APPLICANT:  TAMAOKI, TAIKI
;    APPLICANT:  NAKABAYASHI, HIDEKAZU
;    TITLE OF INVENTION:  IMPROVED METHOD OF EXPRESSING GENES IN
;    TITLE OF INVENTION:  MAMMALIAN CELLS
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BURNS, DOANE, SWECKER & MATHIS
;      STREET:  699 PRINCE STREET
;      CITY:  ALEXANDRIA
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/478,042
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/148,058
;      FILING DATE:  04-NOV-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MOOI, LESLIE A.
;      REGISTRATION NUMBER:  37,047
;      REFERENCE/DOCKET NUMBER:  028722-126
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-854-7400
;      TELEFAX:  415-854-8275
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  13 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-478-042-2


  Query Match           100.0%;  Score 7;  DB 1;  Length 13;
  Best Local Similarity  100.0%;  Pred. No. 6.6e+03;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TACTAAC 7
          |||||||
Db      7 TACTAAC 1


RESULT   7
US-08-645-215-2/c
; Sequence 2, Application US/08645215
; Patent No. 5827686
;  GENERAL INFORMATION:
;    APPLICANT:  TAMAOKI, TAIKI
;    APPLICANT:  NAKABAYASHI, HIDEKAZU
;    TITLE OF INVENTION:  IMPROVED METHOD OF EXPRESSING GENES IN
;    TITLE OF INVENTION:  MAMMALIAN CELLS
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BURNS, DOANE, SWECKER & MATHIS, L.L.P.
;      STREET:  699 PRINCE STREET
;      CITY:  ALEXANDRIA
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
```

```
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/645,215
;       FILING DATE:  13-MAY-1996
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/148,058
;       FILING DATE:  04-NOV-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  MOOI, LESLIE A.
;       REGISTRATION NUMBER:  37,047
;       REFERENCE/DOCKET NUMBER:  028722-135
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  415-854-7400
;       TELEFAX:  415-854-8275
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  13 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
US-08-645-215-2


  Query Match             100.0%;  Score 7;  DB 1;  Length 13;
  Best Local Similarity   100.0%;  Pred. No. 6.6e+03;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TACTAAC 7
           |||||||
Db       7 TACTAAC 1


RESULT   8
US-08-466-604-2/c
; Sequence 2, Application US/08466604
; Patent No. 5843776
;  GENERAL INFORMATION:
;    APPLICANT:  TAMAOKI, TAIKI
;    APPLICANT:  NAKABAYASHI, HIDEKAZU
;    TITLE OF INVENTION:  IMPROVED METHOD OF EXPRESSING GENES IN
;    TITLE OF INVENTION:  MAMMALIAN CELLS
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BURNS, DOANE, SWECKER & MATHIS
;      STREET:  699 PRINCE STREET
;      CITY:  ALEXANDRIA
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/466,604
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/148,058
;      FILING DATE:  04-NOV-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MOOI, LESLIE A.
;      REGISTRATION NUMBER:  37,047
;      REFERENCE/DOCKET NUMBER:  028722-125
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-854-7400
;      TELEFAX:  415-854-8275
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  13 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-466-604-2


  Query Match             100.0%;  Score 7;  DB 2;  Length 13;
  Best Local Similarity   100.0%;  Pred. No. 6.6e+03;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TACTAAC 7
           |||||||
Db       7 TACTAAC 1


RESULT   9
US-08-297-808A-4
; Sequence 4, Application US/08297808A
; Patent No. 5691137
;  GENERAL INFORMATION:
;    APPLICANT:  Rosbash, Michael
;    APPLICANT:  Stutz, Francoise
;    TITLE OF INVENTION:  Methods of Screening Candidate Agents
;    TITLE OF INVENTION:  for Biological Activity Using Yeast Cells
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  Massachusetts
;      COUNTRY:  US
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/297,808A
;      FILING DATE:  30-AUG-1994
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brook, David E.
;      REGISTRATION NUMBER:  22,592
;      REFERENCE/DOCKET NUMBER:  BRU94-01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-861-6240
;      TELEFAX:  617-861-9540
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  14 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-297-808A-4


  Query Match             100.0%;  Score 7;  DB 1;  Length 14;
  Best Local Similarity   71.4%;  Pred. No. 6.6e+03;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TACTAAC 7
           :||:|||
Db       1 UACUAAC 7


RESULT  10
US-09-242-690A-16
; Sequence 16, Application US/09242690A
; Patent No. 6284534
; GENERAL INFORMATION:
```

```
;  APPLICANT: KONDO, KEIJI
;  APPLICANT:  MIURA, YUTAKA
;  TITLE OF INVENTION: YEAST VECTOR AND METHOD OF PRODUCING PROTEINS USING THE
;  TITLE OF INVENTION:  SAME
;  FILE REFERENCE: 049441/0118
;  CURRENT APPLICATION NUMBER: US/09/242,690A
;  CURRENT FILING DATE:  1999-02-23
;  PRIOR APPLICATION NUMBER: PCT/JP97/02924
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: JP 8/241062
;  PRIOR FILING DATE: 1996-08-23
;  NUMBER OF SEQ ID NOS: 66
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
;   LENGTH: 15
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Sequence which
; Patent No. 6284534
;   OTHER INFORMATION: is common to intron
US-09-242-690A-16


  Query Match             100.0%;  Score 7;  DB 4;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 6.5e+03;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TACTAAC 7
          |||||||
Db      7 tactaac 13


RESULT  11
US-07-990-965-3
; Sequence 3, Application US/07990965
; Patent No. 5556954
;  GENERAL INFORMATION:
;    APPLICANT:  Burn, Timothy C.
;    APPLICANT:  Satterthwaite, Anne B.
;    APPLICANT:  Tenen, Daniel G.
;    TITLE OF INVENTION:  Hematopoietic Stem Cell Specific
;    TITLE OF INVENTION:  Gene
;    TITLE OF INVENTION:  Expression
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/990,965
;      FILING DATE:  19921215
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brook, David E.
;      REGISTRATION NUMBER:  22,592
;      REFERENCE/DOCKET NUMBER:  BIH91-03A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617 861 6240
;      TELEFAX:  617 861 9540
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  17 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-07-990-965-3


  Query Match             100.0%;  Score 7;  DB 1;  Length 17;
  Best Local Similarity   100.0%;  Pred. No. 6.5e+03;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TACTAAC 7
          |||||||
Db      3 TACTAAC 9


RESULT  12
US-08-758-306-347
; Sequence 347, Application US/08758306
; Patent No. 5807743
;  GENERAL INFORMATION:
;    APPLICANT:  Stinchcomb, Dan T.
;    APPLICANT:  McSwiggen, James A.
;    TITLE OF INVENTION:  METHOD AND REAGENT FOR THE
;    TITLE OF INVENTION:  TREATMENT OF DISEASES
;    TITLE OF INVENTION:  ASSOCIATED WITH
;    TITLE OF INVENTION:  INTERLEUKIN-2 RECEPTOR
;    TITLE OF INVENTION:  GAMMA-CHAIN EXPRESSION
;    NUMBER OF SEQUENCES:  1379
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  FastSeq Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/758,306
;      FILING DATE:  December 3, 1996
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  212/132
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  347:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  17 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-758-306-347


  Query Match             100.0%;  Score 7;  DB 1;  Length 17;
  Best Local Similarity   71.4%;  Pred. No. 6.5e+03;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TACTAAC 7
          :||:|||
Db      8 UACUAAC 14
```

```
RESULT  13
US-08-758-306-349
; Sequence 349, Application US/08758306
; Patent No. 5807743
;  GENERAL INFORMATION:
;    APPLICANT:  Stinchcomb, Dan T.
;    APPLICANT:  McSwiggen, James A.
;    TITLE OF INVENTION:  METHOD AND REAGENT FOR THE
;    TITLE OF INVENTION:  TREATMENT OF DISEASES
;    TITLE OF INVENTION:  ASSOCIATED WITH
;    TITLE OF INVENTION:  INTERLEUKIN-2 RECEPTOR
;    TITLE OF INVENTION:  GAMMA-CHAIN EXPRESSION
;    NUMBER OF SEQUENCES:  1379
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  FastSeq Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/758,306
;      FILING DATE:  December 3, 1996
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  212/132
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  349:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  17 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-758-306-349


  Query Match             100.0%;  Score 7;  DB 1;  Length 17;
  Best Local Similarity   71.4%;  Pred. No. 6.5e+03;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TACTAAC 7
            :||:|||
Db       5 UACUAAC 11


RESULT  14
US-08-758-306-351
; Sequence 351, Application US/08758306
; Patent No. 5807743
;  GENERAL INFORMATION:
;    APPLICANT:  Stinchcomb, Dan T.
;    APPLICANT:  McSwiggen, James A.
;    TITLE OF INVENTION:  METHOD AND REAGENT FOR THE
;    TITLE OF INVENTION:  TREATMENT OF DISEASES
;    TITLE OF INVENTION:  ASSOCIATED WITH
;    TITLE OF INVENTION:  INTERLEUKIN-2 RECEPTOR
;    TITLE OF INVENTION:  GAMMA-CHAIN EXPRESSION
;    NUMBER OF SEQUENCES:  1379
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  FastSeq Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/758,306
;      FILING DATE:  December 3, 1996
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  212/132
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  351:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  17 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-758-306-351


  Query Match             100.0%;  Score 7;  DB 1;  Length 17;
  Best Local Similarity   71.4%;  Pred. No. 6.5e+03;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TACTAAC 7
            :||:|||
Db       1 UACUAAC 7


RESULT  15
US-08-758-306-1315
; Sequence 1315, Application US/08758306
; Patent No. 5807743
;  GENERAL INFORMATION:
;    APPLICANT:  Stinchcomb, Dan T.
;    APPLICANT:  McSwiggen, James A.
;    TITLE OF INVENTION:  METHOD AND REAGENT FOR THE
;    TITLE OF INVENTION:  TREATMENT OF DISEASES
;    TITLE OF INVENTION:  ASSOCIATED WITH
;    TITLE OF INVENTION:  INTERLEUKIN-2 RECEPTOR
;    TITLE OF INVENTION:  GAMMA-CHAIN EXPRESSION
;    NUMBER OF SEQUENCES:  1379
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
```

```
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  FastSeq Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/758,306
;      FILING DATE:  December 3, 1996
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  212/132
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  1315:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  17 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-758-306-1315


  Query Match            100.0%;  Score 7;  DB 1;  Length 17;
  Best Local Similarity   71.4%;  Pred. No. 6.5e+03;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TACTAAC 7
            :||:|||
Db      11 UACUAAC 17


Search completed: July 21, 2002, 09:47:18
Job time: 11949 sec
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2002, 09:11:03 ; Search time 3274.61 Seconds
(without alignments)
28.852 Million cell updates/sec

Title: US-09-754-014-10_COPY_16_22
Perfect score: 7
Sequence: 1 TACTAAC 7

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	7	100.0	19	12	AZ623493	AZ623493 1M0461M13
	2	7	100.0	19	12	AZ778302	AZ778302 2M0013C02
c	3	7	100.0	19	12	AZ806669	AZ806669 2M0068G19
	4	7	100.0	21	12	AZ812861	AZ812861 2M0079C19
	5	7	100.0	22	12	AZ851620	AZ851620 2M0153M24
	6	7	100.0	23	12	AZ345908	AZ345908 1M0080F22
	7	7	100.0	23	12	AZ476882	AZ476882 1M0296L02
	8	7	100.0	23	12	TA164G12Q	AL473175 T. brucei
	9	7	100.0	24	12	AZ304717	AZ304717 1M0004I20
	10	7	100.0	26	12	AZ364052	AZ364052 1M0110O02
	11	7	100.0	27	12	AZ351430	AZ351430 1M0089C05
c	12	7	100.0	28	12	AZ609297	AZ609297 1M0434B06
	13	7	100.0	30	12	AZ480938	AZ480938 1M0302M09
	14	7	100.0	30	12	AZ591759	AZ591759 1M0402J02
	15	7	100.0	30	12	AZ857764	AZ857764 2M0162M08
c	16	7	100.0	31	12	AZ318049	AZ318049 1M0036F13
	17	7	100.0	31	12	AZ831899	AZ831899 2M0111I22
	18	7	100.0	32	12	AZ387853	AZ387853 1M0147K23
	19	7	100.0	32	12	AZ391582	AZ391582 1M0153I18
c	20	7	100.0	32	12	AZ605009	AZ605009 1M0426E09
	21	7	100.0	32	12	TA253H10P	AL483650 T. brucei
c	22	7	100.0	33	12	AZ769247	AZ769247 1M0569P10
	23	7	100.0	33	12	TA227D09Q	AL479999 T. brucei
	24	7	100.0	33	12	TA364B10P	AL494156 T. brucei
c	25	7	100.0	34	10	BJ040894	BJ040894 BJ040894
c	26	7	100.0	34	12	AZ474616	AZ474616 1M0292B14
	27	7	100.0	34	12	AZ828219	AZ828219 2M0105P01
	28	7	100.0	34	12	TA373F08P	AL496425 T. brucei
	29	7	100.0	35	12	AZ345949	AZ345949 1M0080N23
	30	7	100.0	35	12	TA369E01Q	AL496056 T. brucei
c	31	7	100.0	37	9	AV853613	AV853613 AV853613
	32	7	100.0	37	12	AZ447236	AZ447236 1M0244L18
	33	7	100.0	37	12	AZ465835	AZ465835 1M0276H02
c	34	7	100.0	37	12	AZ490406	AZ490406 1M0323A22
c	35	7	100.0	39	9	AU008671	AU008671 AU008671
	36	7	100.0	39	12	AZ455333	AZ455333 1M0257J08
c	37	7	100.0	40	10	BF381362	BF381362 AsIR0002
	38	7	100.0	40	12	AZ591989	AZ591989 1M0402E06
c	39	7	100.0	41	10	T54451	T54451 yb06h05.r2
	40	7	100.0	41	12	AZ329290	AZ329290 1M0053P22
	41	7	100.0	41	12	AZ504919	AZ504919 1M0345A24
	42	7	100.0	41	12	AZ790809	AZ790809 2M0039B04
	43	7	100.0	42	10	C21089	C21089 HUMGS000260
c	44	7	100.0	42	10	D21033	D21033 HUMGS02016
c	45	7	100.0	42	12	AZ770413	AZ770413 1M0572B01

ALIGNMENTS

RESULT 1
AZ623493/c
LOCUS AZ623493 19 bp DNA linear GSS 13-DEC-2000
DEFINITION 1M0461M13F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0461M13 F, DNA sequence.
ACCESSION AZ623493
VERSION AZ623493.1 GI:11745683
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 19)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0461 row: M column: 13
Seq primer: CGTTGTAAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 19.
FEATURES Location/Qualifiers
source 1. .19
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0461M13"
/clone_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 4 a 3 c 5 g 7 t
ORIGIN

Query Match 100.0%; Score 7; DB 12; Length 19;
Best Local Similarity 100.0%; Pred. No. 2.8e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TACTAAC 7
|||||||
Db 14 TACTAAC 8

RESULT 2
AZ778302
LOCUS AZ778302 19 bp DNA linear GSS 16-FEB-2001
DEFINITION 2M0013C02F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0013C02 F, DNA sequence.
ACCESSION AZ778302
VERSION AZ778302.1 GI:12907800
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 19)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0013 row: C column: 02
Seq primer: CGTTGTAAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 19.
FEATURES Location/Qualifiers
source 1. .19
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0013C02"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 5 a 9 c 0 g 5 t
ORIGIN

Query Match 100.0%; Score 7; DB 12; Length 19;
Best Local Similarity 100.0%; Pred. No. 2.8e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TACTAAC 7
|||||||
Db 8 TACTAAC 14

RESULT 3
AZ806669/c
LOCUS AZ806669 19 bp DNA linear GSS 20-FEB-2001
DEFINITION 2M0068G19R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0068G19 R, DNA sequence.
ACCESSION AZ806669
VERSION AZ806669.1 GI:12970249
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 19)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0068 row: G column: 19
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 19.
FEATURES Location/Qualifiers
source 1. .19
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"

```
                     /clone="UUGC2M0068G19"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT        3 a      2 c      7 g      7 t
ORIGIN


  Query Match             100.0%;  Score 7;  DB 12;  Length 19;
  Best Local Similarity   100.0%;  Pred. No. 2.8e+05;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TACTAAC 7
          |||||||
Db     18 TACTAAC 12


RESULT   4
AZ812861
LOCUS       AZ812861                 21 bp    DNA     linear   GSS 20-FEB-2001
DEFINITION  2M0079C19R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0079C19 R, DNA sequence.
ACCESSION   AZ812861
VERSION     AZ812861.1  GI:12982526
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 21)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0079  row: C  column: 19
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 21.
FEATURES             Location/Qualifiers
     source          1. .21
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0079C19"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT        4 a      9 c      3 g      5 t
ORIGIN


  Query Match             100.0%;  Score 7;  DB 12;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 2.8e+05;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TACTAAC 7
          |||||||
Db      5 TACTAAC 11


RESULT   5
AZ851620
LOCUS       AZ851620                 22 bp    DNA     linear   GSS 21-FEB-2001
DEFINITION  2M0153M24R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0153M24 R, DNA sequence.
ACCESSION   AZ851620
VERSION     AZ851620.1  GI:13037799
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 22)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0153  row: M  column: 24
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 22.
FEATURES             Location/Qualifiers
     source          1. .22
                     /organism="Mus musculus"
```

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0153M24"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 9 a 7 c 0 g 6 t
ORIGIN

Query Match 100.0%; Score 7; DB 12; Length 22;
Best Local Similarity 100.0%; Pred. No. 2.9e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TACTAAC 7
|||||||
Db 6 TACTAAC 12

RESULT 6
AZ345908
LOCUS AZ345908 23 bp DNA linear GSS 29-SEP-2000
DEFINITION 1M0080F22R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0080F22 R, DNA sequence.
ACCESSION AZ345908
VERSION AZ345908.1 GI:10425145
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 23)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0080 row: F column: 22
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 23.
FEATURES Location/Qualifiers
source 1. .23

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0080F22"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 10 a 2 c 1 g 10 t
ORIGIN

Query Match 100.0%; Score 7; DB 12; Length 23;
Best Local Similarity 100.0%; Pred. No. 2.9e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TACTAAC 7
|||||||
Db 13 TACTAAC 19

RESULT 7
AZ476882
LOCUS AZ476882 23 bp DNA linear GSS 04-OCT-2000
DEFINITION 1M0296L02F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0296L02 F, DNA sequence.
ACCESSION AZ476882
VERSION AZ476882.1 GI:10635007
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 23)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0296 row: L column: 02
Seq primer: CGTTGTAAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 23.
FEATURES Location/Qualifiers

```
    source          1. .23
                    /organism="Mus musculus"
                    /strain="C57BL/6J"
                    /db_xref="taxon:10090"
                    /clone="UUGC1M0296L02"
                    /clone_lib="Mouse 10kb plasmid UUGC1M library"
                    /sex="Male"
                    /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                    /note="Vector: PWD42nv; Purified genomic DNA from M.
                    musculus C57BL/6J (male) was obtained  from the Jackson
                    Laboratory Mouse DNA Resource
                    (http://www.jax.org/resources/documents/dnares/). The DNA
                    was hydrodynamically sheared by repeated passage through a
                    0.005 inch orifice at constant velocity. The sheared DNA
                    was blunt end-repaired with T4 DNA polymerase and T4
                    polynucleotide kinase. Adaptor oligonucleotides were
                    ligated to the blunt ends in high  molar excess. The
                    adaptored DNA was purified and size-selected for a 9.5 to
                    10.5 kb range using preparative agarose gel
                    electrophoresis. Vector DNA was prepared from a derivative
                    of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                    inducible derivative of plasmid R1. The vector was ligated
                    with adaptors complementary to the insert adaptors and
                    purified. The sheared, adaptored mouse DNA was annealed to
                    adaptored vector DNA, and transformed into
                    chemically-competent E. coli XL10-Gold (Stratagene) cells
                    and selected for ampicillin resistance."
BASE COUNT        7 a      5 c      2 g      9 t
ORIGIN


  Query Match             100.0%;  Score 7;  DB 12;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 2.9e+05;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TACTAAC 7
          |||||||
Db      4 TACTAAC 10


RESULT   8
TA164G12Q
LOCUS       TA164G12Q                 23 bp    DNA     linear   GSS 13-DEC-2000
DEFINITION  T. brucei sheared genomic DNA clone 164g12, reverse sequence,
            genomic survey sequence.
ACCESSION   AL473175
VERSION     AL473175.1  GI:11838448
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei.
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 23)
  AUTHORS   Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
            Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
            Melville,S.E., Rajandream,M.A. and Barrell,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
            project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
            Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
            nh1@sanger.ac.uk
COMMENT     Constructed at the Institute for Genomic Research (TIGR),
            Rockville, MD. Genomic DNA isolated from a cloned population of
            Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
            to give a tight size distribution (
            4 kb).  The v + i method used for the library construction is
            described in detail in Smith, H. and Venter, J.C. (Making small
            insert libraries for whole genome shotgun sequencing projects.  In
            Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
            Barrell, Oxford University Press, 1999).
            Email: nelsayed@tigr.org
            Details of T. brucei sequencing at the Sanger Centre are available
            at http://www.sanger.ac.uk/Projects/T_brucei/.
FEATURES            Location/Qualifiers
     source         1. .23
                    /organism="Trypanosoma brucei"
                    /strain="TREU927"
                    /db_xref="taxon:5691"
                    /clone="164g12"
BASE COUNT       11 a      5 c      1 g      6 t
ORIGIN


  Query Match             100.0%;  Score 7;  DB 12;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 2.9e+05;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TACTAAC 7
          |||||||
Db      2 TACTAAC 8


RESULT   9
AZ304717
LOCUS       AZ304717                  24 bp    DNA     linear   GSS 29-SEP-2000
DEFINITION  1M0004I20R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0004I20 R, DNA sequence.
ACCESSION   AZ304717
VERSION     AZ304717.1  GI:10341011
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 24)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0004  row: I  column: 20
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 24.
FEATURES            Location/Qualifiers
     source         1. .24
                    /organism="Mus musculus"
                    /strain="C57BL/6J"
                    /db_xref="taxon:10090"
                    /clone="UUGC1M0004I20"
                    /clone_lib="Mouse 10kb plasmid UUGC1M library"
                    /sex="Male"
                    /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                    /note="Vector: PWD42nv; Purified genomic DNA from M.
                    musculus C57BL/6J (male) was obtained  from the Jackson
                    Laboratory Mouse DNA Resource
                    (http://www.jax.org/resources/documents/dnares/). The DNA
                    was hydrodynamically sheared by repeated passage through a
                    0.005 inch orifice at constant velocity. The sheared DNA
                    was blunt end-repaired with T4 DNA polymerase and T4
                    polynucleotide kinase. Adaptor oligonucleotides were
                    ligated to the blunt ends in high  molar excess. The
                    adaptored DNA was purified and size-selected for a 9.5 to
                    10.5 kb range using preparative agarose gel
                    electrophoresis. Vector DNA was prepared from a derivative
```

of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 7 a 8 c 3 g 6 t
ORIGIN

Query Match 100.0%; Score 7; DB 12; Length 24;
Best Local Similarity 100.0%; Pred. No. 2.9e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TACTAAC 7
|||||||
Db 2 TACTAAC 8

RESULT 10
AZ364052
LOCUS AZ364052 26 bp DNA linear GSS 02-OCT-2000
DEFINITION 1M0110O02F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0110O02 F, DNA sequence.
ACCESSION AZ364052
VERSION AZ364052.1 GI:10477752
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 26)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0110 row: O column: 02
Seq primer: CGTTGTAAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 26.
FEATURES Location/Qualifiers
source 1. .26
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0110O02"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 10 a 10 c 0 g 6 t
ORIGIN

Query Match 100.0%; Score 7; DB 12; Length 26;
Best Local Similarity 100.0%; Pred. No. 2.9e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TACTAAC 7
|||||||
Db 16 TACTAAC 22

RESULT 11
AZ351430
LOCUS AZ351430 27 bp DNA linear GSS 29-SEP-2000
DEFINITION 1M0089C05R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0089C05 R, DNA sequence.
ACCESSION AZ351430
VERSION AZ351430.1 GI:10430667
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 27)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0089 row: C column: 05
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 27.
FEATURES Location/Qualifiers
source 1. .27
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0089C05"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackscn Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 10 a 7 c 2 g 8 t
ORIGIN

Query Match 100.0%; Score 7; DB 12; Length 27;
Best Local Similarity 100.0%; Pred. No. 2.9e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TACTAAC 7
|||||||
Db 10 TACTAAC 16

RESULT 12
AZ609297/c
LOCUS AZ609297 28 bp DNA linear GSS 13-DEC-2000
DEFINITION 1M0434B06F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0434B06 F, DNA sequence.
ACCESSION AZ609297
VERSION AZ609297.1 GI:11731487
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 28)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0434 row: B column: 06
Seq primer: CGTTGTAAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 28.
FEATURES Location/Qualifiers
source 1. .28
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0434B06"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 9 a 1 c 9 g 9 t
ORIGIN

Query Match 100.0%; Score 7; DB 12; Length 28;
Best Local Similarity 100.0%; Pred. No. 2.9e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TACTAAC 7
|||||||
Db 22 TACTAAC 16

RESULT 13
AZ480938
LOCUS AZ480938 30 bp DNA linear GSS 04-OCT-2000
DEFINITION 1M0302M09R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0302M09 R, DNA sequence.
ACCESSION AZ480938
VERSION AZ480938.1 GI:10641919
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 30)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0302 row: M column: 09
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 30.
FEATURES Location/Qualifiers
source 1. .30
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0302M09"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 9 a 9 c 3 g 9 t
ORIGIN

Query Match 100.0%; Score 7; DB 12; Length 30;
Best Local Similarity 100.0%; Pred. No. 2.9e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TACTAAC 7
|||||||
Db 18 TACTAAC 24

RESULT 14
AZ591759
LOCUS AZ591759 30 bp DNA linear GSS 13-DEC-2000
DEFINITION 1M0402J02F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0402J02 F, DNA sequence.
ACCESSION AZ591759
VERSION AZ591759.1 GI:11713949
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 30)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0402 row: J column: 02
Seq primer: CGTTGTAAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 30.
FEATURES Location/Qualifiers
source 1. .30
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0402J02"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 8 a 6 c 8 g 8 t
ORIGIN

Query Match 100.0%; Score 7; DB 12; Length 30;
Best Local Similarity 100.0%; Pred. No. 2.9e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TACTAAC 7
|||||||
Db 13 TACTAAC 19

RESULT 15
AZ857764
LOCUS AZ857764 30 bp DNA linear GSS 21-FEB-2001
DEFINITION 2M0162M08R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0162M08 R, DNA sequence.
ACCESSION AZ857764
VERSION AZ857764.1 GI:13050236
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 30)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0162 row: M column: 08
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 30.
FEATURES Location/Qualifiers
source 1. .30
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0162M08"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 6 a 4 c 8 g 12 t
ORIGIN

Query Match 100.0%; Score 7; DB 12; Length 30;
Best Local Similarity 100.0%; Pred. No. 2.9e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 TACTAAC 7
           |||||||
Db       6 TACTAAC 12
```

Search completed: July 21, 2002, 09:11:05
Job time: 10381 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2002, 09:45:15 ; Search time 2038.31 Seconds
(without alignments)
215.599 Million cell updates/sec

Title: US-09-754-014-10_COPY_25_45
Perfect score: 21
Sequence: 1 TTCTTTTTTTCTCTTCACAGG 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	21	100.0	45	6	BD007073	BD007073 IL-12 gen
2	21	100.0	45	6	BD007083	BD007083 Gene expr
3	21	100.0	3589	6	AX249943	AX249943 Sequence
4	21	100.0	3609	6	AX249944	AX249944 Sequence
5	21	100.0	4276	6	AX249946	AX249946 Sequence
6	21	100.0	4496	6	AX249945	AX249945 Sequence
c 7	20	95.2	362	1	AF302194	AF302194 Streptoco
c 8	19.4	92.4	154761	2	AC098692	AC098692 Homo sapi
c 9	19.4	92.4	169434	9	AL359205	AL359205 Human DNA
10	19.4	92.4	171588	2	AC023198	AC023198 Homo sapi
11	19.4	92.4	203668	2	AC084411	AC084411 Mus muscu
12	19.4	92.4	208618	2	AC019313	AC019313 Homo sapi
13	19	90.5	162522	2	AC092824	AC092824 Homo sapi
c 14	18.4	87.6	2283	5	AF033189	AF033189 Gallus ga
c 15	18.4	87.6	3382	6	AX082205	AX082205 Sequence
c 16	18.4	87.6	3589	9	AF126484	AF126484 Homo sapi
c 17	18.4	87.6	12534	3	AF099002	AF099002 Caenorhab
c 18	18.4	87.6	35804	3	CBRG20K13	AC084524 Caenorhab
c 19	18.4	87.6	40214	3	U32305	U32305 Caenorhabdi
20	18.4	87.6	47972	9	AL391843	AL391843 Human DNA
c 21	18.4	87.6	64797	2	AC102695	AC102695 Mus muscu
c 22	18.4	87.6	122704	9	AC026406	AC026406 Homo sapi
c 23	18.4	87.6	137342	9	AL392048	AL392048 Human DNA
24	18.4	87.6	140678	9	AC079922	AC079922 Homo sapi
25	18.4	87.6	148380	2	AC012609	AC012609 Homo sapi
26	18.4	87.6	149386	2	AC040997	AC040997 Homo sapi
27	18.4	87.6	152869	2	AC020890	AC020890 Homo sapi
28	18.4	87.6	152966	9	AL359713	AL359713 Human DNA
c 29	18.4	87.6	158309	2	AC097170	AC097170 Rattus no
30	18.4	87.6	171247	2	AC108016	AC108016 Homo sapi
c 31	18.4	87.6	180388	9	HUMRETBLAS	L11910 Human retin
c 32	18.4	87.6	181835	9	AC026124	AC026124 Homo sapi
c 33	18.4	87.6	187737	2	AC021026	AC021026 Homo sapi
c 34	18.4	87.6	195859	14	AF281817	AF281817 Tupaia he
c 35	18.4	87.6	203269	2	AC080187	AC080187 Homo sapi
c 36	18.4	87.6	215532	2	AC098478	AC098478 Homo sapi
c 37	18	85.7	125045	2	AC094470	AC094470 Rattus no
38	18	85.7	142331	2	AF307157	AF307157 Homo sapi
c 39	18	85.7	151136	9	HSA189K21	AL121893 Human DNA
40	18	85.7	153168	2	AC087045	AC087045 Homo sapi
c 41	18	85.7	158999	2	AC083843	AC083843 Homo sapi
42	18	85.7	168804	9	AL354920	AL354920 Human DNA
c 43	18	85.7	170028	2	AC095693	AC095693 Rattus no
44	18	85.7	222794	2	AC099462	AC099462 Rattus no
c 45	18	85.7	321003	2	PFMAL4P3	AL035476 Plasmodiu

ALIGNMENTS

RESULT 1
BD007073
LOCUS BD007073 45 bp DNA linear PAT 31-JAN-2002
DEFINITION IL-12 gene expression and delivery systems and uses.
ACCESSION BD007073
VERSION BD007073.1 GI:18635444
KEYWORDS JP 2001503257-A/4.
SOURCE unidentified.
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 45)
AUTHORS Nodosutoromu,J., Freemark,B. and Dishupande,D.
TITLE IL-12 gene expression and delivery systems and uses
JOURNAL Patent: JP 2001503257-A 4 13-MAR-2001;
BARENTISU INC,SYNTEX INC
COMMENT OS Unidentified
PN JP 2001503257-A/4
PD 13-MAR-2001
PF 10-OCT-1997 JP 1998519514
PR 18-OCT-1996 US 60/028676
PI JEFF NODOSUTOROMU,BLUCE FREEMARK,DIPA DISHUPANDE PC
C12N15/09,A61K31/711,A61K38/00,A61K47/18,A61K47/28,A61K48/00, PC
A61P11/06,

```
            PC   C12N15/00,A61K37/02
            CC   Strandedness: Single;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers
            FT   source          1. .45
            FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .45
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        8 a      10 c       8 g      19 t
ORIGIN


  Query Match             100.0%;  Score 21;  DB 6;  Length 45;
  Best Local Similarity   100.0%;  Pred. No. 87;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TTCTTTTTTTCTCTTCACAGG 21
          |||||||||||||||||||||
Db     25 TTCTTTTTTTCTCTTCACAGG 45


RESULT   2
BD007083
LOCUS       BD007083                 45 bp    DNA     linear   PAT 31-JAN-2002
DEFINITION  Gene expression and delivery systems and uses.
ACCESSION   BD007083
VERSION     BD007083.1  GI:18635454
KEYWORDS    JP 2001503258-A/8.
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 45)
  AUTHORS   Nodosutoromu,J., Freemark,B. and Dishupande,D.
  TITLE     Gene expression and delivery systems and uses
  JOURNAL   Patent: JP 2001503258-A 8 13-MAR-2001;
            BARENTISU INC
COMMENT     OS   Unidentified
            PN   JP 2001503258-A/8
            PD   13-MAR-2001
            PF   10-OCT-1997 JP 1998519520
            PR   18-OCT-1996 US    60/028687
            PI   JEFF NODOSUTOROMU,BLUCE FREEMARK,DIPA DISHUPANDE PC
            C12N15/09,A61K48/00,A61P11/06,A61P35/00,A61P43/00//C07K14/54, PC
            C12N15/00
            CC   Strandedness: Single;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers
            FT   source          1. .45
            FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .45
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        8 a      10 c       8 g      19 t
ORIGIN


  Query Match             100.0%;  Score 21;  DB 6;  Length 45;
  Best Local Similarity   100.0%;  Pred. No. 87;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TTCTTTTTTTCTCTTCACAGG 21
          |||||||||||||||||||||
Db     25 TTCTTTTTTTCTCTTCACAGG 45


RESULT   3
AX249943
LOCUS       AX249943                 3589 bp    DNA     linear   PAT 28-SEP-2001
DEFINITION  Sequence 1 from Patent WO0166149.
ACCESSION   AX249943
VERSION     AX249943.1  GI:15864429
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 3589)
  AUTHORS   Fewell,J.G., Maclaughlin,F., Smith,L.C., Nicol,F. and Rolland,A.
  TITLE     Nucleic acid formulations for gene delivery and methods of use
  JOURNAL   Patent: WO 0166149-A 1 13-SEP-2001;
            Valentis, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .3589
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="Expression plasmid pIF0921 encoding for human
                     interferon alpha (7 68) __. (1334)."
     CDS             768. .1334
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAC88666.1"
                     /db_xref="GI:15864430"
                     /translation="MALTFALLVALLVLSCKSSCSVGCDLPQTHSLGSRRTLMLLAQM
                     RRISLFSCLKNRHDFGFPQEEFGNQFQKAETIPVLHEMIQQIFNLFSTKDSSAAWDET
                     LLDKFYTELYQQLNDLEACVIQGVGVTETPLMKEDSILAVRKYFQRITLYLKEKKYSP
                     CAWEVVRAEIMRSFSLSTNLQESLRSKE"
BASE COUNT      833 a    983 c    932 g    841 t
ORIGIN


  Query Match             100.0%;  Score 21;  DB 6;  Length 3589;
  Best Local Similarity   100.0%;  Pred. No. 58;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TTCTTTTTTTCTCTTCACAGG 21
          |||||||||||||||||||||
Db    742 TTCTTTTTTTCTCTTCACAGG 762


RESULT   4
AX249944
LOCUS       AX249944                 3609 bp    DNA     linear   PAT 28-SEP-2001
DEFINITION  Sequence 2 from Patent WO0166149.
ACCESSION   AX249944
VERSION     AX249944.1  GI:15864431
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 3609)
  AUTHORS   Fewell,J.G., Maclaughlin,F., Smith,L.C., Nicol,F. and Rolland,A.
  TITLE     Nucleic acid formulations for gene delivery and methods of use
  JOURNAL   Patent: WO 0166149-A 2 13-SEP-2001;
            Valentis, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .3609
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="Expression plasmid pEP1403 encoding for mouse
                     erythropoietin (801 ) __. (1379)"
     CDS             801. .1379
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAC88667.1"
                     /db_xref="GI:15864432"
                     /translation="MGVPERPTLLLLLSLLLIPLGLPVLCAPPRLICDSRVLERYILE
                     AKEAENVTMGCAEGPRLSENITVPDTKVNFYAWKRMEVEEQAIEVWQGLSLLSEAILQ
                     AQALLANSSQPPETLQLHIDKAISGLRSLTSLLRVLGAQKELMSPPDTTPPAPLRTLT
                     VDTFCKLFRVYANFLRGKLKLYTGEVCRRGDR"
BASE COUNT      811 a   1031 c    953 g    814 t
ORIGIN
```

```
  Query Match            100.0%;  Score 21;  DB 6;  Length 3609;
  Best Local Similarity  100.0%;  Pred. No. 58;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TTCTTTTTTTCTCTTCACAGG 21
          |||||||||||||||||||||
Db    767 TTCTTTTTTTCTCTTCACAGG 787


RESULT   5
AX249946
LOCUS       AX249946                4276 bp    DNA     linear   PAT 28-SEP-2001
DEFINITION  Sequence 4 from Patent WO0166149.
ACCESSION   AX249946
VERSION     AX249946.1  GI:15864435
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 4276)
  AUTHORS   Fewell,J.G., Maclaughlin,F., Smith,L.C., Nicol,F. and Rolland,A.
  TITLE     Nucleic acid formulations for gene delivery and methods of use
  JOURNAL   Patent: WO 0166149-A 4 13-SEP-2001;
            Valentis, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .4276
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="Expression plasmid pFN1645 having codon optimized
                     sequence encodi ng for human coagulation factor IX (786)
                     __. (2171)."
     CDS             786. .2171
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAC88669.1"
                     /db_xref="GI:15864436"
                     /translation="MQRVNMIMAESPGLITICLLGYLLSAECTVFLDHENANKILNRP
                     KRYNSGKLEEFVQGNLERECMEEKCSFEEAREVFENTERTTEFWKQYVDGDQCESNPC
                     LNGGSCKDDINSYECWCPFGFEGKNCELDVTCNIKNGRCEQFCKNSADNKVVCSCTEG
                     YRLAENQKSCEPAVPFPCGRVSVSQTSKLTRAEAVFPDVDYVNSTEAETILDNITQST
                     QSFNDFTRVVGGEDAKPGQFPWQVVLNGKVDAFCGGSIVNEKWIVTAAHCVETGVKIT
                     VVAGEHNIEETEHTEQKRNVIRIIPHHNYNAAINKYNHDIALLELDEPLVLNSYVTPI
                     CIADKEYTNIFLKFGSGYVSGWGRVFHKGRSALVLQYLRVPLVDRATCLRSTKFTIYN
                     NMFCAGFHEGGRDSCQGDSGGPHVTEVEGTSFLTGIISWGEECAMKGKYGIYTKVSRY
                     VNWIKEKTKLT"
BASE COUNT     1059 a    1092 c    1120 g    1005 t
ORIGIN


  Query Match            100.0%;  Score 21;  DB 6;  Length 4276;
  Best Local Similarity  100.0%;  Pred. No. 57;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TTCTTTTTTTCTCTTCACAGG 21
          |||||||||||||||||||||
Db    749 TTCTTTTTTTCTCTTCACAGG 769


RESULT   6
AX249945
LOCUS       AX249945                4496 bp    DNA     linear   PAT 28-SEP-2001
DEFINITION  Sequence 3 from Patent WO0166149.
ACCESSION   AX249945
VERSION     AX249945.1  GI:15864433
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 4496)
  AUTHORS   Fewell,J.G., Maclaughlin,F., Smith,L.C., Nicol,F. and Rolland,A.
  TITLE     Nucleic acid formulations for gene delivery and methods of use
  JOURNAL   Patent: WO 0166149-A 3 13-SEP-2001;
            Valentis, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .4496
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="Expression plasmid pFN0945 having natural sequence
                     encoding human coagulation factor IX"
     CDS             782. .2167
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAC88668.1"
                     /db_xref="GI:15864434"
                     /translation="MQRVNMIMAESPGLITICLLGYLLSAECTVFLDHENANKILNRP
                     KRYNSGKLEEFVQGNLERECMEEKCSFEEAREVFENTERTTEFWKQYVDGDQCESNPC
                     LNGGSCKDDINSYECWCPFGFEGKNCELDVTCNIKNGRCEQFCKNSADNKVVCSCTEG
                     YRLAENQKSCEPAVPFPCGRVSVSQTSKLTRAETVFPDVDYVNSTEAETILDNITQST
                     QSFNDFTRVVGGEDAKPGQFPWQVVLNGKVDAFCGGSIVNEKWIVTAAHCVETGVKIT
                     VVAGEHNIEETEHTEQKRNVIRIIPHHNYNAAINKYNHDIALLELDEPLVLNSYVTPI
                     CIADKEYTNIFLKFGSGYVSGWGRVFHKGRSALVLQYLRVPLVDRATCLRSTKFTIYN
                     NMFCAGFHEGGRDSCQGDSGGPHVTEVEGTSFLTGIISWGEECAMKGKYGIYTKVSRY
                     VNWIKEKTKLT"
BASE COUNT     1127 a    1119 c    1147 g    1103 t
ORIGIN


  Query Match            100.0%;  Score 21;  DB 6;  Length 4496;
  Best Local Similarity  100.0%;  Pred. No. 56;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TTCTTTTTTTCTCTTCACAGG 21
          |||||||||||||||||||||
Db    748 TTCTTTTTTTCTCTTCACAGG 768


RESULT   7
AF302194/c
LOCUS       AF302194                362 bp    DNA     linear   BCT 02-OCT-2001
DEFINITION  Streptococcus suis clone iri7 iron-restricted induced promoter
            region.
ACCESSION   AF302194
VERSION     AF302194.1  GI:15824351
KEYWORDS    .
SOURCE      Streptococcus suis.
  ORGANISM  Streptococcus suis
            Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (bases 1 to 362)
  AUTHORS   Smith,H.E., Buijs,H., de Vries R,R., Wisselink,H.J.,
            Stockhofe-Zurwieden,N. and Smits,M.A.
  TITLE     Environmentally regulated genes of Streptococcus suis:
            identification by the use of iron-restricted conditions in vitro
            and by experimental infection of piglets
  JOURNAL   Microbiology 147 (Pt 2), 271-280 (2001)
  MEDLINE   21097266
   PUBMED   11158344
REFERENCE   2  (bases 1 to 362)
  AUTHORS   Smith,H.E., Buijs,H., de Vries,R., Wisselink,H.J.,
            Stockhofe-Zurwieden,N. and Smits,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-2000) Department of Bacteriology, Institute for
            Animal Science and Health, P.O. Box 65, Lelystad 8200 AB, The
            Netherlands
FEATURES             Location/Qualifiers
     source          1. .362
                     /organism="Streptococcus suis"
                     /db_xref="taxon:1307"
                     /clone="iri7"
     misc_feature    1. .362
                     /note="includes iron-restricted induced promoter"
BASE COUNT      124 a      56 c      74 g     100 t       8 others
ORIGIN
```

```
  Query Match            95.2%;  Score 20;  DB 1;  Length 362;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TTCTTTTTTTCTCTTCACAG 20
          ||||||||||||||||||||
Db    129 TTCTTTTTTTCTCTTCACAG 110


RESULT   8
AC098692/c
LOCUS       AC098692              154761 bp    DNA     linear   HTG 30-OCT-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-60K15, WORKING DRAFT SEQUENCE,
            7 unordered pieces.
ACCESSION   AC098692
VERSION     AC098692.1  GI:16519529
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 154761)
  AUTHORS   Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 154761)
  AUTHORS   Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-OCT-2001) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
COMMENT     -------------- Genome Center
                Center: University of Washington Genome Center
                Center Code: UWGC
                Web site: http://www.genome.washington.edu
                Contact: uwgchtgs@u.washington.edu
            -------------- Project Information
                Center project name: chr-1
                Center clone name: RP11-60K15 (sc0443)
            ------------- Summary Statistics
                Sequencing vector: plasmid; L08752; 100% of reads
                Chemistry: Dye-terminator ET; 84% of reads
                Chemistry: Dye-terminator Big Dye; 16% of reads
                Assembly program: Phrap; version 0.990319
                Consensus quality: 150156 bases at least Q40
                Consensus quality: 152633 bases at least Q30
                Consensus quality: 153757 bases at least Q20
                Insert size: 154161; sum-of-contigs
                Quality coverage: 11.9x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2732: contig of 2732 bp in length
            *     2733     2832: gap of unknown length
            *     2833     6152: contig of 3320 bp in length
            *     6153     6252: gap of unknown length
            *     6253    14109: contig of 7857 bp in length
            *    14110    14209: gap of unknown length
            *    14210    23378: contig of 9169 bp in length
            *    23379    23478: gap of unknown length
            *    23479    39562: contig of 16084 bp in length
            *    39563    39662: gap of unknown length
            *    39663    75978: contig of 36316 bp in length
            *    75979    76078: gap of unknown length
            *    76079   154761: contig of 78683 bp in length.
FEATURES             Location/Qualifiers
     source          1. .154761
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-60K15"
                     /clone_lib="RPCI human BAC library 11"
     misc_feature    1. .2732
                     /note="assembly_name:Contig15"
     misc_feature    2833. .6152
                     /note="assembly_name:Contig16"
     misc_feature    6253. .14109
                     /note="assembly_name:Contig17"
     misc_feature    14210. .23378
                     /note="assembly_name:Contig18"
     misc_feature    23479. .39562
                     /note="assembly_name:Contig19"
     misc_feature    39663. .75978
                     /note="assembly_name:Contig20"
     misc_feature    76079. .154761
                     /note="assembly_name:Contig21"
BASE COUNT    50562 a  26801 c  25413 g  51382 t    603 others
ORIGIN


  Query Match            92.4%;  Score 19.4;  DB 2;  Length 154761;
  Best Local Similarity   95.2%;  Pred. No. 1.9e+02;
  Matches   20;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TTCTTTTTTTCTCTTCACAGG 21
            || |||||||||||||||||
Db 102134 TTTTTTTTTTCTCTTCACAGG 102114


RESULT   9
AL359205/c
LOCUS       AL359205              169434 bp    DNA     linear   PRI 13-DEC-2000
DEFINITION  Human DNA sequence from clone RP11-345N16 on chromosome 1, complete
            sequence.
ACCESSION   AL359205
VERSION     AL359205.15  GI:11863412
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169434)
  AUTHORS   Williams,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Dec 15, 2000 this sequence version replaced gi:11691506.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows.  An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated repeat sequence elements.  Where the sequence is
            ambiguous, there is an annotation using the 'unsure' feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP11-345N16 is from the library RPCI-11.2 constructed by the group
```

```
          of Pieter de Jong. For further details see
          http://www.chori.org/bacpac/home.htm
          VECTOR: pBACe3.6
          This sequence is the entire insert of clone RP11-345N16.
FEATURES             Location/Qualifiers
     source          1. .169434
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-345N16"
                     /clone_lib="RPCI-11.2"
     repeat_region   174. .235
                     /note="31 copies 2 mer ca 90% conserved"
     repeat_region   1140. .1222
                     /note="MER77 repeat: matches 237. .320 of consensus"
     repeat_region   1471. .1582
                     /note="MER77 repeat: matches 527. .644 of consensus"
     repeat_region   2576. .2996
                     /note="MLT1C repeat: matches 1. .466 of consensus"
     repeat_region   3093. .3480
                     /note="LTR7 repeat: matches 1. .450 of consensus"
     repeat_region   3769. .3998
                     /note="L2 repeat: matches 2004. .2248 of consensus"
     repeat_region   4267. .4405
                     /note="L2 repeat: matches 2553. .2690 of consensus"
     repeat_region   4985. .5016
                     /note="16 copies 2 mer tt 84% conserved"
     repeat_region   5721. .6028
                     /note="AluY repeat: matches 1. .309 of consensus"
     repeat_region   9207. .9501
                     /note="AluSq repeat: matches 1. .275 of consensus"
     repeat_region   9772. .10079
                     /note="AluSc repeat: matches 6. .298 of consensus"
     repeat_region   10559. .10769
                     /note="MIR repeat: matches 20. .248 of consensus"
     repeat_region   11527. .11757
                     /note="L2 repeat: matches 2482. .2710 of consensus"
     repeat_region   11870. .12423
                     /note="L2 repeat: matches 1731. .2340 of consensus"
     repeat_region   12572. .12637
                     /note="L2 repeat: matches 2647. .2710 of consensus"
     repeat_region   12838. .12936
                     /note="L2 repeat: matches 2645. .2750 of consensus"
     repeat_region   13946. .14244
                     /note="L1PA13 repeat: matches 5858. .6156 of consensus"
     repeat_region   14914. .14955
                     /note="21 copies 2 mer ac 78% conserved"
     repeat_region   14959. .15066
                     /note="MER6 repeat: matches 753. .861 of consensus"
     repeat_region   15068. .17642
                     /note="L1P repeat: matches 2. .2537 of consensus"
     repeat_region   17647. .17706
                     /note="Alu repeat: matches 239. .298 of consensus"
     repeat_region   17745. .18452
                     /note="MER6 repeat: matches 2. .763 of consensus"
     repeat_region   18454. .18489
                     /note="18 copies 2 mer tg 97% conserved"
     repeat_region   18620. .18742
                     /note="MLT1J repeat: matches 122. .251 of consensus"
     repeat_region   19184. .19351
                     /note="MER69A repeat: matches 4. .176 of consensus"
     repeat_region   19604. .19665
                     /note="31 copies 2 mer ca 67% conserved"
     repeat_region   21165. .21305
                     /note="L2 repeat: matches 1702. .1845 of consensus"
     repeat_region   21433. .21777
                     /note="MLT1A1 repeat: matches 10. .360 of consensus"
     repeat_region   21822. .22009
                     /note="L2 repeat: matches 1331. .1515 of consensus"
     repeat_region   22634. .22696
                     /note="MIR repeat: matches 200. .262 of consensus"
     repeat_region   22798. .23109
                     /note="AluSg repeat: matches 1. .310 of consensus"
     repeat_region   23921. .24249
                     /note="HERVL repeat: matches 5085. .5427 of consensus"
     repeat_region   25484. .25840
                     /note="MER46B repeat: matches 2. .236 of consensus"
     repeat_region   28006. .28390
                     /note="MLT1B repeat: matches 1. .389 of consensus"
     repeat_region   29010. .29172
                     /note="MIR repeat: matches 67. .245 of consensus"
     repeat_region   29334. .29539
                     /note="MIR repeat: matches 15. .241 of consensus"
     repeat_region   29762. .30706
                     /note="L1PA5 repeat: matches 3361. .4322 of consensus"
     repeat_region   30702. .32480
                     /note="L1PA5 repeat: matches 4355. .6143 of consensus"
     repeat_region   32929. .32997
                     /note="L1M4 repeat: matches 4645. .4713 of consensus"
     repeat_region   33003. .33537
                     /note="MLT1E repeat: matches 1. .560 of consensus"
     repeat_region   35344. .35381
                     /note="19 copies 2 mer ca 97% conserved"
     repeat_region   35769. .36156
                     /note="LTR37A repeat: matches 46. .421 of consensus"
     repeat_region   36991. .37288
                     /note="MER39 repeat: matches 55. .354 of consensus"
     repeat_region   37945. .37982
                     /note="L1PA12 repeat: matches -1422. .-1386 of consensus"
     repeat_region   38073. .38376
                     /note="AluSq repeat: matches 1. .312 of consensus"
     repeat_region   38420. .38611
                     /note="L1PA12 repeat: matches -1285. .-1099 of consensus"
     repeat_region   38761. .41603
                     /note="L1PA12 repeat: matches -822. .3083 of consensus"
     repeat_region   42994. .43343
                     /note="MER57A repeat: matches 12. .396 of consensus"
     repeat_region   44944. .45187
                     /note="MSTB repeat: matches 3. .271 of consensus"
     repeat_region   45195. .48302
                     /note="L1PA7 repeat: matches 3000. .6135 of consensus"
     repeat_region   48325. .48535
                     /note="MLT1A1 repeat: matches 144. .357 of consensus"
     repeat_region   48535. .48881
                     /note="MLT1B repeat: matches 43. .390 of consensus"
     repeat_region   49529. .50398
                     /note="L2 repeat: matches 1781. .2687 of consensus"
     repeat_region   50634. .50934
                     /note="AluSg repeat: matches 5. .305 of consensus"
     repeat_region   52188. .52482
                     /note="L1MB8 repeat: matches 5892. .6166 of consensus"
     repeat_region   53076. .53181
                     /note="MIR repeat: matches 85. .191 of consensus"
     repeat_region   53653. .53703
                     /note="L1MB7 repeat: matches 6114. .6164 of consensus"
     repeat_region   53697. .54260
                     /note="L1MC1 repeat: matches 4744. .5329 of consensus"
     repeat_region   54264. .54433
                     /note="FRAM repeat: matches 1. .176 of consensus"
     repeat_region   54440. .55424
                     /note="L1MC1 repeat: matches 5341. .6314 of consensus"
     repeat_region   55451. .55538
                     /note="44 copies 2 mer at 77% conserved"
     repeat_region   55544. .55703
                     /note="80 copies 2 mer at 57% conserved"
     repeat_region   55843. .55894
                     /note="26 copies 2 mer ca 78% conserved"
     repeat_region   55896. .55929
                     /note="17 copies 2 mer ta 97% conserved"
     repeat_region   55944. .56522
                     /note="L1MB6 repeat: matches 5555. .6127 of consensus"
     repeat_region   56522. .56581
                     /note="L1MA5 repeat: matches 3906. .3964 of consensus"
     repeat_region   56582. .56890
                     /note="AluY repeat: matches 3. .310 of consensus"
     repeat_region   56891. .57507
```

```
                     /note="L1MA5 repeat: matches 3964. .4566 of consensus"
     repeat_region   57512. .57861
                     /note="THE1C repeat: matches 15. .371 of consensus"
     repeat_region   57862. .59586
                     /note="L1MA5 repeat: matches 4561. .6300 of consensus"
     repeat_region   59595. .59783
                     /note="L1M4 repeat: matches 5396. .5568 of consensus"
     repeat_region   59780. .60717
                     /note="L1M4 repeat: matches 4309. .5287 of consensus"
     repeat_region   60724. .60892
                     /note="L1M4 repeat: matches 3243. .3420 of consensus"
     repeat_region   60894. .60989
                     /note="48 copies 2 mer tt 66% conserved"
     repeat_region   61123. .61328
                     /note="L1M4 repeat: matches 2870. .3091 of consensus"
     repeat_region   61330. .61426

  Query Match             92.4%;  Score 19.4;  DB 9;  Length 169434;
  Best Local Similarity   95.2%;  Pred. No. 1.9e+02;
  Matches   20;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 TTCTTTTTTTCTCTTCACAGG 21
           || ||||||||||||||||||
Db 111932 TTTTTTTTTTCTCTTCACAGG 111912


RESULT  10
AC023198
LOCUS       AC023198              171588 bp    DNA     linear   HTG 28-MAR-2000
DEFINITION  Homo sapiens chromosome 1 clone RP11-345N16 map 1, WORKING DRAFT
            SEQUENCE, 18 unordered pieces.
ACCESSION   AC023198
VERSION     AC023198.2  GI:7331456
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 171588)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 1, clone RP11-345N16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 171588)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 28, 2000 this sequence version replaced gi:6957780.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5962
                Center clone name: 345_N_16
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 161125 bases at least Q40
                Consensus quality: 165418 bases at least Q30
                Consensus quality: 167739 bases at least Q20
                Insert size: 176000; agarose-fp
                Insert size: 169888; sum-of-contigs
                Quality coverage: 5.0 in Q20 bases; agarose-fp
                Quality coverage: 5.2 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 18 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     2304: contig of 2304 bp in length
            *    2305 2404: gap of      100 bp
            *    2405     4339: contig of 1935 bp in length
            *    4340 4439: gap of      100 bp
            *    4440     7773: contig of 3334 bp in length
            *    7774 7873: gap of      100 bp
            *    7874    11812: contig of 3939 bp in length
            *   11813 11912: gap of      100 bp
            *   11913    15783: contig of 3871 bp in length
            *   15784 15883: gap of      100 bp
            *   15884    20493: contig of 4610 bp in length
            *   20494 20593: gap of      100 bp
            *   20594    24877: contig of 4284 bp in length
            *   24878 24977: gap of      100 bp
            *   24978    29978: contig of 5001 bp in length
            *   29979 30078: gap of      100 bp
            *   30079    36535: contig of 6457 bp in length
            *   36536 36635: gap of      100 bp
            *   36636    44045: contig of 7410 bp in length
            *   44046 44145: gap of      100 bp
            *   44146    52402: contig of 8257 bp in length
            *   52403 52502: gap of      100 bp
            *   52503    62821: contig of 10319 bp in length
            *   62822 62921: gap of      100 bp
            *   62922    72744: contig of 9823 bp in length
            *   72745 72844: gap of      100 bp
            *   72845    88300: contig of 15456 bp in length
            *   88301 88400: gap of      100 bp
            *   88401   104799: contig of 16399 bp in length
            *  104800 104899: gap of      100 bp
            *  104900   123814: contig of 18915 bp in length
            *  123815 123914: gap of      100 bp
            *  123915   146837: contig of 22923 bp in length
            *  146838 146937: gap of      100 bp
            *  146938   171588: contig of 24651 bp in length.
FEATURES             Location/Qualifiers
     source          1. .171588
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="1"
                     /clone="RP11-345N16"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .2304
                     /note="assembly_fragment"
     misc_feature    2405. .4339
                     /note="assembly_fragment"
     misc_feature    4440. .7773
                     /note="assembly_fragment"
     misc_feature    7874. .11812
                     /note="assembly_fragment"
```

```
     misc_feature    11913. .15783
                     /note="assembly_fragment"
     misc_feature    15884. .20493
                     /note="assembly_fragment"
     misc_feature    20594. .24877
                     /note="assembly_fragment"
     misc_feature    24978. .29978
                     /note="assembly_fragment"
     misc_feature    30079. .36535
                     /note="assembly_fragment"
     misc_feature    36636. .44045
                     /note="assembly_fragment"
     misc_feature    44146. .52402
                     /note="assembly_fragment"
     misc_feature    52503. .62821
                     /note="assembly_fragment"
     misc_feature    62922. .72744
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    72845. .88300
                     /note="assembly_fragment"
     misc_feature    88401. .104799
                     /note="assembly_fragment"
     misc_feature    104900. .123814
                     /note="assembly_fragment"
     misc_feature    123915. .146837
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    146938. .171588
                     /note="assembly_fragment"
BASE COUNT    54344 a  29344 c  29364 g  56833 t   1703 others
ORIGIN


  Query Match             92.4%;  Score 19.4;  DB 2;  Length 171588;
  Best Local Similarity   95.2%;  Pred. No. 1.9e+02;
  Matches   20;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TTCTTTTTTTCTCTTCACAGG 21
          || ||||||||||||||||||
Db 149939 TTTTTTTTTTCTCTTCACAGG 149959


RESULT  11
AC084411
LOCUS       AC084411               203668 bp    DNA     linear   HTG 01-NOV-2000
DEFINITION  Mus musculus clone RP23-125M20, WORKING DRAFT SEQUENCE, 32
            unordered pieces.
ACCESSION   AC084411
VERSION     AC084411.1  GI:11067260
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 203668)
  AUTHORS   McCombie,W.R., Baker,J.P., Bahret,A., Bal,H., Balija,V.,
            Dedhia,N.N., de la Bastide,M., Huang,E.N., King,L., Kirchoff,K.A.,
            Miller,B., Nascimento,L.U., O'Shaughnessy,A.L., Preston,R.R.,
            Rodriguez,S., Santos,L., Shah,R.S., Spiegel,L.A., Toth,K., Vil,M.D.
            and Zutavern,T.
  TITLE     Mouse Genomic Sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 203668)
  AUTHORS   McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-NOV-2000) Lita Annenberg Hazen Genome Sequencing
            Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
            Harbor, NY 11724, USA
COMMENT     ------------------- Genome Center
            Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor
            Laboratory
            Center code: CSHL
            Web site: http://www.cshl.org/genseq
            Contact:  mccombie@cshl.org
            ------------------- Project Information
            Center project name: RP23-125M20
            Center clone name: RP23-125M20
            -------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 32 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    21298: contig of 21298 bp in length
            *   21299    21390: gap of unknown length
            *   21391    37719: contig of 16329 bp in length
            *   37720    37811: gap of unknown length
            *   37812    52285: contig of 14474 bp in length
            *   52286    52377: gap of unknown length
            *   52378    66315: contig of 13938 bp in length
            *   66316    66407: gap of unknown length
            *   66408    79620: contig of 13213 bp in length
            *   79621    79712: gap of unknown length
            *   79713    92050: contig of 12338 bp in length
            *   92051    92142: gap of unknown length
            *   92143   102157: contig of 10015 bp in length
            *  102158   102249: gap of unknown length
            *  102250   110283: contig of 8034 bp in length
            *  110284   110375: gap of unknown length
            *  110376   117524: contig of 7149 bp in length
            *  117525   117616: gap of unknown length
            *  117617   124463: contig of 6847 bp in length
            *  124464   124555: gap of unknown length
            *  124556   130800: contig of 6245 bp in length
            *  130801   130892: gap of unknown length
            *  130893   136994: contig of 6102 bp in length
            *  136995   137086: gap of unknown length
            *  137087   143118: contig of 6032 bp in length
            *  143119   143210: gap of unknown length
            *  143211   148264: contig of 5054 bp in length
            *  148265   148355: gap of unknown length
            *  148356   153298: contig of 4943 bp in length
            *  153299   153389: gap of unknown length
            *  153390   158314: contig of 4925 bp in length
            *  158315   158405: gap of unknown length
            *  158406   162746: contig of 4341 bp in length
            *  162747   162837: gap of unknown length
            *  162838   167078: contig of 4241 bp in length
            *  167079   167169: gap of unknown length
            *  167170   170577: contig of 3408 bp in length
            *  170578   170668: gap of unknown length
            *  170669   173666: contig of 2998 bp in length
            *  173667   173757: gap of unknown length
            *  173758   176695: contig of 2938 bp in length
            *  176696   176786: gap of unknown length
            *  176787   179696: contig of 2910 bp in length
            *  179697   179787: gap of unknown length
            *  179788   182656: contig of 2869 bp in length
            *  182657   182747: gap of unknown length
            *  182748   185462: contig of 2715 bp in length
            *  185463   185553: gap of unknown length
            *  185554   188209: contig of 2656 bp in length
            *  188210   188300: gap of unknown length
            *  188301   190826: contig of 2526 bp in length
            *  190827   190917: gap of unknown length
            *  190918   193443: contig of 2526 bp in length
            *  193444   193534: gap of unknown length
            *  193535   195903: contig of 2369 bp in length
            *  195904   195994: gap of unknown length
            *  195995   198332: contig of 2338 bp in length
```

```
            *   198333   198423: gap of unknown length
            *   198424   200537: contig of 2114 bp in length
            *   200538   200628: gap of unknown length
            *   200629   202638: contig of 2010 bp in length
            *   202639   202729: gap of unknown length
            *   202730   203668: contig of 939 bp in length.
FEATURES             Location/Qualifiers
     source          1. .203668
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP23-125M20"
                     /clone_lib="RPCI-23"
BASE COUNT    53265 a  46398 c  45935 g  54709 t   3361 others
ORIGIN


  Query Match             92.4%;  Score 19.4;  DB 2;  Length 203668;
  Best Local Similarity   95.2%;  Pred. No. 1.8e+02;
  Matches   20;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 TTCTTTTTTTCTCTTCACAGG 21
           ||||||||||||||||  ||||
Db 179609 TTCTTTTTTTCTCTTCCCAGG 179629


RESULT  12
AC019313
LOCUS       AC019313               208618 bp    DNA     linear   HTG 17-MAR-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-119P12 map 18, WORKING DRAFT
            SEQUENCE, 49 unordered pieces.
ACCESSION   AC019313
VERSION     AC019313.3  GI:7259732
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 208618)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone RP11-119P12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 208618)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-DEC-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 17, 2000 this sequence version replaced gi:6984442.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5386
                Center clone name: 119_P_12
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 161318 bases at least Q40
                Consensus quality: 181693 bases at least Q30
                Consensus quality: 194050 bases at least Q20
                Insert size: 203818; sum-of-contigs
                Quality coverage: 3.2 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 49 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1052: contig of 1052 bp in length
            *     1053 1152: gap of      100 bp
            *     1153     2486: contig of 1334 bp in length
            *     2487 2586: gap of      100 bp
            *     2587     3719: contig of 1133 bp in length
            *     3720 3819: gap of      100 bp
            *     3820     3915: contig of 96 bp in length
            *     3916 4015: gap of      100 bp
            *     4016     5118: contig of 1103 bp in length
            *     5119 5218: gap of      100 bp
            *     5219     6275: contig of 1057 bp in length
            *     6276 6375: gap of      100 bp
            *     6376     7721: contig of 1346 bp in length
            *     7722 7821: gap of      100 bp
            *     7822     8956: contig of 1135 bp in length
            *     8957 9056: gap of      100 bp
            *     9057    10726: contig of 1670 bp in length
            *    10727 10826: gap of      100 bp
            *    10827    12216: contig of 1390 bp in length
            *    12217 12316: gap of      100 bp
            *    12317    13744: contig of 1428 bp in length
            *    13745 13844: gap of      100 bp
            *    13845    15089: contig of 1245 bp in length
            *    15090 15189: gap of      100 bp
            *    15190    16739: contig of 1550 bp in length
            *    16740 16839: gap of      100 bp
            *    16840    17920: contig of 1081 bp in length
            *    17921 18020: gap of      100 bp
            *    18021    19143: contig of 1123 bp in length
            *    19144 19243: gap of      100 bp
            *    19244    20489: contig of 1246 bp in length
            *    20490 20589: gap of      100 bp
            *    20590    21851: contig of 1262 bp in length
            *    21852 21951: gap of      100 bp
            *    21952    23358: contig of 1407 bp in length
            *    23359 23458: gap of      100 bp
            *    23459    25234: contig of 1776 bp in length
            *    25235 25334: gap of      100 bp
            *    25335    27169: contig of 1835 bp in length
            *    27170 27269: gap of      100 bp
            *    27270    29801: contig of 2532 bp in length
            *    29802 29901: gap of      100 bp
            *    29902    32268: contig of 2367 bp in length
            *    32269 32368: gap of      100 bp
            *    32369    34415: contig of 2047 bp in length
            *    34416 34515: gap of      100 bp
            *    34516    36552: contig of 2037 bp in length
            *    36553 36652: gap of      100 bp
            *    36653    39620: contig of 2968 bp in length
            *    39621 39720: gap of      100 bp
            *    39721    42451: contig of 2731 bp in length
            *    42452 42551: gap of      100 bp
            *    42552    46217: contig of 3666 bp in length
            *    46218 46317: gap of      100 bp
```

```
            *    46318    48687: contig of 2370 bp in length
            *    48688 48787: gap of      100 bp
            *    48788    52204: contig of 3417 bp in length
            *    52205 52304: gap of      100 bp
            *    52305    55124: contig of 2820 bp in length
            *    55125 55224: gap of      100 bp
            *    55225    59236: contig of 4012 bp in length
            *    59237 59336: gap of      100 bp
            *    59337    63427: contig of 4091 bp in length
            *    63428 63527: gap of      100 bp
            *    63528    68235: contig of 4708 bp in length
            *    68236 68335: gap of      100 bp
            *    68336    74336: contig of 6001 bp in length
            *    74337 74436: gap of      100 bp
            *    74437    79648: contig of 5212 bp in length
            *    79649 79748: gap of      100 bp
            *    79749    86527: contig of 6779 bp in length
            *    86528 86627: gap of      100 bp
            *    86628    93092: contig of 6465 bp in length
            *    93093 93192: gap of      100 bp
            *    93193    97736: contig of 4544 bp in length
            *    97737 97836: gap of      100 bp
            *    97837   103589: contig of 5753 bp in length
            *   103590 103689: gap of      100 bp
            *   103690   111389: contig of 7700 bp in length
            *   111390 111489: gap of      100 bp
            *   111490   117615: contig of 6126 bp in length
            *   117616 117715: gap of      100 bp
            *   117716   126146: contig of 8431 bp in length
            *   126147 126246: gap of      100 bp
            *   126247   133762: contig of 7516 bp in length
            *   133763 133862: gap of      100 bp
            *   133863   143183: contig of 9321 bp in length
            *   143184 143283: gap of      100 bp
            *   143284   154009: contig of 10726 bp in length
            *   154010 154109: gap of      100 bp
            *   154110   164224: contig of 10115 bp in length
            *   164225 164324: gap of      100 bp
            *   164325   176540: contig of 12216 bp in length
            *   176541 176640: gap of      100 bp
            *   176641   192181: contig of 15541 bp in length
            *   192182 192281: gap of      100 bp
            *   192282   208618: contig of 16337 bp in length.
FEATURES             Location/Qualifiers
     source          1. .208618
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18"
                     /clone="RP11-119P12"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1052
                     /note="assembly_fragment"
     misc_feature    1153. .2486
                     /note="assembly_fragment"
     misc_feature    2587. .3719
                     /note="assembly_fragment"
     misc_feature    3820. .3915
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    4016. .5118
                     /note="assembly_fragment"
     misc_feature    5219. .6275
                     /note="assembly_fragment"
     misc_feature    6376. .7721
                     /note="assembly_fragment"
     misc_feature    7822. .8956
                     /note="assembly_fragment"
     misc_feature    9057. .10726
                     /note="assembly_fragment"
     misc_feature    10827. .12216
                     /note="assembly_fragment"
     misc_feature    12317. .13744
                     /note="assembly_fragment"
     misc_feature    13845. .15089
                     /note="assembly_fragment"
     misc_feature    15190. .16739
                     /note="assembly_fragment"
     misc_feature    16840. .17920
                     /note="assembly_fragment"

  Query Match             92.4%;  Score 19.4;  DB 2;  Length 208618;
  Best Local Similarity   95.2%;  Pred. No. 1.8e+02;
  Matches   20;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 TTCTTTTTTTCTCTTCACAGG 21
           |||||||||||||||||| |
Db   61291 TTCTTTTTTTCTCTTCACACG 61311


RESULT  13
AC092824
LOCUS       AC092824              162522 bp    DNA     linear   HTG 02-DEC-2001
DEFINITION  Homo sapiens chromosome 12p clone RP11-158N24, WORKING DRAFT
            SEQUENCE, 2 unordered pieces.
ACCESSION   AC092824
VERSION     AC092824.10  GI:17223142
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 162522)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 162522)
```

```
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUL-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Dec 1, 2001 this sequence version replaced gi:17136092.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HDKG
                Center clone name: RP11-158N24
            -------------- Summary Statistics
                Sequencing vector: Plasmid; M77789
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 162690 bases at least Q40
                Consensus quality: 162922 bases at least Q30
                Consensus quality: 163111 bases at least Q20
                Estimated insert size: 162375; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 10.2x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   108667: contig of 108667 bp in length
            *  108668   108767: gap of unknown length
            *  108768   162522: contig of 53755 bp in length.
FEATURES             Location/Qualifiers
     source          1. .162522
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="12p"
                     /clone="RP11-158N24"
BASE COUNT    46500 a  35361 c  32754 g  47807 t    100 others
ORIGIN


  Query Match             90.5%;  Score 19;  DB 2;  Length 162522;
  Best Local Similarity   100.0%;  Pred. No. 2.7e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TTCTTTTTTTCTCTTCACA 19
          ||||||||||||||||||
Db  86548 TTCTTTTTTTCTCTTCACA 86566


RESULT  14
AF033189/c
LOCUS       AF033189               2283 bp    mRNA    linear   VRT 13-DEC-1997
DEFINITION  Gallus gallus sulfotransferase mRNA, complete cds.
ACCESSION   AF033189
VERSION     AF033189.1  GI:2687359
KEYWORDS    .
SOURCE      chicken.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 2283)
  AUTHORS   Cao,H., Agarwal,S. and Burnside,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1997) Animal and Food Science, University of
            Delaware, 40 Townsend Hall, Newark, DE 19350, USA
FEATURES             Location/Qualifiers
     source          1. .2283
                     /organism="Gallus gallus"
                     /strain="domesticus"
                     /db_xref="taxon:9031"
                     /tissue_type="liver"
     CDS             71. .1009
                     /codon_start=1
                     /product="sulfotransferase"
                     /protein_id="AAB88818.1"
                     /db_xref="GI:2687360"
                     /translation="MEKSRKKFIDVIDKAIVIGNAMDRDELLFSYKGVLYPVALCSPE
                     VFRAMESFEARSDDVILAGYPKSGTNWVGQILSDLVATFEKERLEEKSVNDEELEEFP
                     YLEIGDTEKYERMKKLPSRRVILTHLSPEKLPKSIFKNKAKILLLIRNPKDIATSFFH
                     FSNRWSALPSYETWDDFFIAFMTEKMPWGSYFNYLSEWNKYAADENVMTITYEELKEN
                     QTLGVKNIASFFGISLTGEELRSVIERSSFQSMKENSLKTHGALGSMLFRKGGVSDWK
                     NLFNEEQNEKMDKVFEERIARTKLGTKLKYEVYCKA"
BASE COUNT      749 a    401 c    447 g    686 t
ORIGIN


  Query Match             87.6%;  Score 18.4;  DB 5;  Length 2283;
  Best Local Similarity   95.0%;  Pred. No. 7.2e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TTCTTTTTTTCTCTTCACAG 20
          || |||||||||||||||||
Db   2281 TTTTTTTTTTCTCTTCACAG 2262


RESULT  15
AX082205/c
LOCUS       AX082205               3382 bp    DNA     linear   PAT 27-FEB-2001
DEFINITION  Sequence 7 from Patent WO0100826.
ACCESSION   AX082205
VERSION     AX082205.1  GI:13170989
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3382)
  AUTHORS   Bertin,J.
  TITLE     Novel molecules of the card-related protein family and uses thereof
  JOURNAL   Patent: WO 0100826-A 7 04-JAN-2001;
            Millennium Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .3382
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             245. .3106
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC33156.1"
                     /db_xref="GI:13170990"
                     /translation="MEEQGHSEMEIIPSESHPHIQLLKSNRELLVTHIRNTQCLVDNL
                     LKNDYFSAEDAEIVCACPTQPDKVRKILDLVQSKGEEVSEFFLYLLQQLADAYVDLRP
                     WLLEIGFSPSLLTQSKVVVNTDPVSRYTQQLRHHLGRDSKFVLCYAQKEELLLEEIYM
                     DTIMELVGFSNESLGSLNSLACLLDHTTGILNEQGETIFILGDAGVGKSMLLQRLQSL
                     WATGRLDAGVKFFFHFRCRMFSCFKESDRLCLQDLLFKHYCYPERDPEEVFAFLLRFP
                     HVALFTFDGLDELHSDLDLSRVPDSSCPWEPAHPLVLLANLLSGKLLKGASKLLTART
                     GIEVPRQFLRKKVLLRGFSPSHLRAYARRMFPERALQDRLLSQLEANPNLCSLCSVPL
                     FCWIIFRCFQHFRAAFEGSPQLPDCTMTLTDVFLLVTEVHLNRMQPSSLVQRNTRSPV
                     ETLHAGRDTLCSLGQVAHRGMEKSLFVFTQEEVQASGLQERDMQLGFLRALPELGPGG
                     DQQSYEFFHLTLQAFFTAFFLVLDDRVGTQELLRFFQEWMPPAGAATTSCYPPFLPFQ
                     CLQGSGPAREDLFKNKDHFQFTNLFLCGLLSKAKQKLLRHLVPAAALRRKRKALWAHL
                     FSSLRGYLKSLPRVQVESFNQVQAMPTFIWMLRCIYETQSQKVGQLAARGICANYLKL
                     TYCNACSADCSALSFVLHHFPKRLALDLDNNNLNDYGVRELQPCFSRLTVLRLSVNQI
                     TDGGVKVLSEELTKYKIVTYLGLYNNQITDVGARYVTKILDECKGLTHLKLGKNKITS
                     EGGKYLALAVKNSKSISEVGMWGNQVGDEGAKAFAEALRNHPSLTTLSLASNGISTEG
                     GKSLARALQQNTSLEILWLTQNELNDEVAESLAEMLKVNQTLKHLWLIQNQITAKGTA
                     QLADALQSNTGITEICLNGNLIKPEEAKVYEDEKRIICF"
```

BASE COUNT 775 a 975 c 933 g 693 t 6 others
ORIGIN

Query Match 87.6%; Score 18.4; DB 6; Length 3382;
Best Local Similarity 95.0%; Pred. No. 6.9e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1 TTCTTTTTTTCTCTTCACAG 20
           || |||||||||||||||||
Db    3371 TTTTTTTTTTCTCTTCACAG 3352
```

Search completed: July 21, 2002, 09:45:33
Job time: 12324 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2002, 09:55:19 ; Search time 467.25 Seconds
(without alignments)
77.165 Million cell updates/sec

Title: US-09-754-014-10_COPY_25_45
Perfect score: 21
Sequence: 1 TTCTTTTTTTCTCTTCACAGG 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	21	100.0	30	20	AAZ40413	3' splice site seq
	2	21	100.0	45	19	AAV07276	Plasmid pIN0773 In
	3	21	100.0	45	21	AAZ50395	Synthetic intron,
	4	21	100.0	3589	20	AAZ40418	Plasmid pIF0921 co
	5	21	100.0	3589	22	AAI70084	Plasmid pIF0921 en
	6	21	100.0	3609	22	AAI70085	Plasmid pEP1403 en
	7	21	100.0	4276	22	AAI70087	Codon optimised pl
	8	21	100.0	4496	22	AAI70086	Plasmid pFN0945 en
	9	21	100.0	5686	20	AAZ40415	Plasmid pIN1143 co
	10	21	100.0	5966	20	AAZ40417	Plasmid pIN0961 co
c	11	18.4	87.6	1462	20	AAZ09247	Human CARD-4L part
c	12	18.4	87.6	3382	22	AAF30002	Human CARD-4L (lon
c	13	17.8	84.8	366	22	AAI87537	Human polynucleoti
c	14	17.8	84.8	3810	22	AAK87602	Human immune/haema
	15	17.8	84.8	4098	16	AAT43682	Medium chain-speci
	16	17.8	84.8	39746	23	ABL13398	Drosophila melanog
	17	17.8	84.8	161425	22	AAH02340	Human AKAP10 gene
	18	17.8	84.8	162025	22	AAH02339	Human AKAP10 gene
	19	17.4	82.9	495	21	AAC75463	Human ORFX ORF1018
	20	17.4	82.9	754	22	AAI97059	Human neuroblastom
c	21	17.4	82.9	1930	20	AAX02108	Mouse FEN-1 cDNA.
c	22	17.4	82.9	2033	20	AAX02111	Human FEN-1 genomi
c	23	17.4	82.9	6749	22	AAS46526	Tumour suppressor
c	24	17.4	82.9	19480	22	AAK80384	Human immune/haema
c	25	17.4	82.9	19481	22	AAK80383	Human immune/haema
c	26	17.4	82.9	19932	17	AAT46159	CagI locus. Helic
c	27	17.4	82.9	30013	22	AAL36932	Human musculoskele
	28	17.4	82.9	30013	22	AAS41960	Genomic sequence #
	29	17.4	82.9	96583	21	AAF22297	BAC containing rep
c	30	17	81.0	1225	15	AAQ73396	CviJI ORF1 coding
c	31	17	81.0	5496	15	AAQ73395	CviJI coding seque
c	32	16.8	80.0	304	14	AAQ60333	Human brain Expres
c	33	16.8	80.0	386	16	AAT26283	Human gene signatu
	34	16.8	80.0	414	22	AAK63511	Human immune/haema
	35	16.8	80.0	414	22	AAK64652	Human immune/haema
c	36	16.8	80.0	420	22	AAL18660	Human breast cance
	37	16.8	80.0	1215	22	AAI66509	Pig caspase coding
c	38	16.8	80.0	1216	21	AAC51370	Arabidopsis thalia
c	39	16.8	80.0	1254	21	AAC52518	Arabidopsis thalia
	40	16.8	80.0	1374	22	AAK71327	Human immune/haema
	41	16.8	80.0	1374	22	AAK71357	Human immune/haema
	42	16.8	80.0	1419	22	AAS40333	DNA encoding human
	43	16.8	80.0	1419	22	AAL03933	Human reproductive
	44	16.8	80.0	1529	22	AAS40335	DNA encoding human
	45	16.8	80.0	1529	22	AAS40341	DNA encoding human

ALIGNMENTS

RESULT 1
AAZ40413
ID AAZ40413 standard; DNA; 30 BP.
XX
AC AAZ40413;
XX
DT 15-FEB-2000 (first entry)
XX
DE 3' splice site sequence for interferon-alpha plasmid.
XX
KW Wild type; human; interferon-alpha; plasmid; cytomegalovirus; CMV;
KW promoter; growth hormone; untranslated region; UTR; mammal; disease;
KW cancer; intron; ss.
XX
OS Synthetic.
XX
PN WO9947678-A2.
XX
PD 23-SEP-1999.
XX
PF 12-MAR-1999; 99WO-US05394.
XX
PR 19-MAR-1998; 98US-0078654.
XX
PA (GENE-) GENEMEDICINE INC.
XX
PI Nordstrom J, Pericle F, Rolland A, Ralston R;
XX
DR WPI; 1999-562116/47.
XX
PT New plasmids containing an interferon-alpha coding sequence, used for
PT the treatment of a mammalian condition or disease, particularly cancer

```
PT   -
XX
PS   Disclosure; Page 31; 137pp; English.
XX
CC   The invention relates to a novel plasmid comprising a cytomegalovirus
CC   (CMV) promoter transcriptionally linked with an interferon alpha
CC   (IFN-alpha) coding sequence, and a growth hormone 3'-untranslated
CC   region (UTR).  Sequences AAZ40412 and AAZ40413 represent synthetic 5' and
CC   3' splice donor and acceptor sites respectively for generating a
CC   synthetic intron to be inserted into the plasmid of the invention. The
CC   plasmids can be used for treating a mammalian condition or disease,
CC   particularly cancer.
XX
SQ   Sequence 30 BP; 5 A; 7 C; 4 G; 14 T; 0 other;


  Query Match             100.0%;  Score 21;  DB 20;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TTCTTTTTTTCTCTTCACAGG 21
          |||||||||||||||||||||
Db     10 ttctttttttctcttcacagg 30


RESULT   2
AAV07276
ID   AAV07276 standard; DNA; 45 BP.
XX
AC   AAV07276;
XX
DT   25-SEP-1998  (first entry)
XX
DE   Plasmid pIN0773 Intron.
XX
KW   IL-12 subunit; expression construct; treatment; asthma; microbial
KW   infection; viral infection; cancer; Human; Interleukin; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1..9
FT                   /*tag=  a
FT                   /note= "5'splice site"
FT   misc_feature    15..16
FT                   /*tag=  b
FT                   /note= "Unspecified sequence of 77 bp not given"
FT   misc_feature    16..22
FT                   /*tag=  c
FT                   /note= "Branch point"
FT   misc_feature    25..44
FT                   /*tag=  d
FT                   /note= "3' splice site"
XX
PN   WO9817689-A2.
XX
PD   30-APR-1998.
XX
PF   10-OCT-1997;   97WO-US18779.
XX
PR   18-OCT-1996;   96US-0028676.
XX
PA   (GENE-) GENEMEDICINE INC.
XX
PI   Deshpande D,  Freimark B,  Nordstrom J;
XX
DR   WPI; 1998-261428/23.
XX
PT   Constructs for expression of interleukin-12 sub-units  - are used
PT   for delivery of IL-12 sub-units for treating e.g. asthma, microbial
PT   or viral infections and certain cancers
XX
PS   Disclosure; Page 26; 80pp; English.
XX
CC   The synthetic intron was designed for highly efficient and accurate RNA
CC   splicing.  The intron was used in the plasmid pIN0773 which can provide
CC   for efficient expression of IL-12 subunits.  The products can be used for
CC   the treatment of asthma, microbial and viral infections and certain
CC   cancers.
XX
SQ   Sequence 45 BP; 8 A; 10 C; 8 G; 19 T; 0 other;


  Query Match             100.0%;  Score 21;  DB 19;  Length 45;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TTCTTTTTTTCTCTTCACAGG 21
          |||||||||||||||||||||
Db     25 ttctttttttctcttcacagg 45


RESULT   3
AAZ50395
ID   AAZ50395 standard; DNA; 45 BP.
XX
AC   AAZ50395;
XX
DT   18-MAY-2000  (first entry)
XX
DE   Synthetic intron, OPTIVS8.
XX
KW   Synthetic intron; OPTIVS8; expression plasmid; anti-angiogenic agent;
KW   cancer; translation; gene expression; RNA splicing; transfection;
KW   tumour activity; solid tumour; lung metastatic tumour; cytostatic;
KW   gene therapy; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    3..4
FT                   /*tag=  a
FT                   /label= 5'_splice_site
FT                   /note= "Corresponds to BbsI cleavage site"
FT   misc_feature    15..16
FT                   /*tag=  b
FT                   /note= "There are 77 residues between C15 and T16 that
FT                   are not shown in the specification"
FT   misc_feature    44..44
FT                   /*tag=  c
FT                   /label= 3'splice_site
FT                   /note= "Corresponds to EarI cleavage site"
XX
PN   WO200006759-A2.
XX
PD   10-FEB-2000.
XX
PF   20-JUL-1999;   99WO-US16388.
XX
PR   27-JUL-1998;   98US-0094375.
XX
PA   (VALE-) VALENTIS INC.
XX
PI   Min W,  Szymanski P,  Mehrens D,  Ralston R,  Sullivan S;
XX
DR   WPI; 2000-183133/16.
XX
PT   Plasmids comprising tissue specific transcription elements linked to an
PT   anti-angiogenic gene is useful transfection of cells and treatment of,
PT   e.g. cancer  -
XX
PS   Disclosure; Page 34; 103pp; English.
XX
CC   The present sequence is a synthetic intron, OPTIVS8 used in the
CC   construction of the expression plasmid incorporating an anti-angiogenic
CC   agent for the treatment of mammalian diseases, especially cancer. This
```

```
CC   intron was designed for effective RNA splicing and increased gene
CC   expression. The plasmids can be used for (in vivo) transfection of a
CC   cell in situ in order to modulate tumour activity. Anti-angiogenic gene
CC   inhibits growth of solid tumour and lung metastatic tumours by
CC   intravenous or intramuscular delivery.
XX
SQ   Sequence 45 BP; 8 A; 10 C; 8 G; 19 T; 0 other;


  Query Match             100.0%;  Score 21;  DB 21;  Length 45;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TTCTTTTTTTCTCTTCACAGG 21
          |||||||||||||||||||||
Db     25 ttctttttttctcttcacagg 45


RESULT   4
AAZ40418
ID   AAZ40418 standard; DNA; 3589 BP.
XX
AC   AAZ40418;
XX
DT   15-FEB-2000  (first entry)
XX
DE   Plasmid pIF0921 containing human IFN-a sequence.
XX
KW   Wild type; human; interferon-alpha; plasmid; cytomegalovirus; CMV;
KW   promoter; growth hormone; untranslated region; UTR; mammal; disease;
KW   cancer; intron; ss.
XX
OS   Synthetic.
XX
PN   WO9947678-A2.
XX
PD   23-SEP-1999.
XX
PF   12-MAR-1999;   99WO-US05394.
XX
PR   19-MAR-1998;   98US-0078654.
XX
PA   (GENE-) GENEMEDICINE INC.
XX
PI   Nordstrom J,  Pericle F,  Rolland A,  Ralston R;
XX
DR   WPI; 1999-562116/47.
XX
PT   New plasmids containing an interferon-alpha coding sequence, used for
PT   the treatment of a mammalian condition or disease, particularly cancer
PT   -
XX
PS   Disclosure; Fig 6; 137pp; English.
XX
CC   The invention relates to a novel plasmid comprising a cytomegalovirus
CC   (CMV) promoter transcriptionally linked with an interferon alpha
CC   (IFN-alpha) coding sequence, and a growth hormone 3'-untranslated
CC   region (UTR).  This sequence represents the plasmid pIF0921 which
CC   contains the human interferon alpha (IFN-a) gene. The plasmids can be
CC   used for treating a mammalian condition or disease, particularly cancer.
XX
SQ   Sequence 3589 BP; 832 A; 983 C; 933 G; 841 T; 0 other;


  Query Match             100.0%;  Score 21;  DB 20;  Length 3589;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TTCTTTTTTTCTCTTCACAGG 21
          |||||||||||||||||||||
Db    742 ttctttttttctcttcacagg 762


RESULT   5
AAI70084
ID   AAI70084 standard; DNA; 3589 BP.
XX
AC   AAI70084;
XX
DT   21-DEC-2001  (first entry)
XX
DE   Plasmid pIF0921 encoding human interferon-alpha.
XX
KW   Plasmid pIF0921; interferon-alpha; cytokine; human; gene delivery;
KW   immune disorder; gene therapy; vaccine; ds.
XX
OS   Chimeric - Homo sapiens.
OS   Chimeric - human cytomegalovirus.
XX
FH   Key             Location/Qualifiers
FT   CDS             768..1334
FT                   /*tag= a
FT                   /product= "human interferon-alpha"
XX
PN   WO200166149-A2.
XX
PD   13-SEP-2001.
XX
PF   02-MAR-2001; 2001WO-US06953.
XX
PR   03-MAR-2000; 2000US-187236P.
PR   16-JAN-2001; 2001US-261751P.
XX
PA   (VALE-) VALENTIS INC.
XX
PI   Fewell JG,  Maclaughlin F,  Smith LC,  Nicol F,  Rolland A;
XX
DR   WPI; 2001-638995/73.
XX
PT   Nucleic acid formulation for gene delivery to a muscle or tumour tissue
PT   to treat cancer, or infectious disease in a mammal, comprises a nucleic
PT   acid and non-encapsulating anionic polymer such as poly-L-glutamate   -
XX
PS   Example 5; Page 88-90; 98pp; English.
XX
CC   The present sequence is that of expression plasmid pIF0921, which
CC   encodes human interferon-alpha (IFN-alpha).  The IFN-alpha coding
CC   sequence was inserted into the Valentis plasmid backbone containing
CC   a 107 bp 5' untranslated region, a 117 bp synthetic intron, the
CC   human growth hormone polyadenylation signal, a pUC12 origin of
CC   replication and a kanamycin resistance gene, such that the
CC   IFN-alpha gene was driven by the cytomegalovirus enhancer/promoter.
CC   The resulting plasmid, pIF0921, was formulated with poly-L-glutamate
CC   to produce a gene delivery vehicle, which was intramuscularly
CC   injected into mice; both legs were electroporated with caliper
CC   electrodes.  A significant enhancement of human IFN-alpha
CC   expression in CD-1 mice was observed.  This is an example of a
CC   method designed for non-viral plasmid-based gene therapy.  In this
CC   method, a nucleic acid is formulated with a non-encapsulating
CC   anionic polymer, such as (biodegradable) poly-L-glutamate, which
CC   enhances transfection of the nucleic acid into muscle or tumour
CC   tissues, with or without electroporation, and which also stabilises
CC   the nucleic acid during storage.  The formulations allow for
CC   vaccination and treatment of muscle disorders and serum protein
CC   deficiencies, as well as cancer and infections.  In the case of
CC   IFN-alpha gene delivery, it may be used to treat a disease
CC   characterised by dysregulation of the immune system.
XX
SQ   Sequence 3589 BP; 833 A; 983 C; 932 G; 841 T; 0 other;


  Query Match             100.0%;  Score 21;  DB 22;  Length 3589;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TTCTTTTTTTCTCTTCACAGG 21
```

```
          ||||||||||||||||||||
Db    742 ttctttttttctcttcacagg 762


RESULT   6
AAI70085
ID   AAI70085 standard; DNA; 3609 BP.
XX
AC   AAI70085;
XX
DT   21-DEC-2001  (first entry)
XX
DE   Plasmid pEP1403 encoding mouse erythropoietin.
XX
KW   Plasmid pEP1403; erythropoietin; mouse; gene delivery;
KW   anaemia; gene therapy; vaccine; ds.
XX
OS   Chimeric - Homo sapiens.
OS   Chimeric - human cytomegalovirus.
XX
FH   Key             Location/Qualifiers
FT   CDS             801..1379
FT                   /*tag= a
FT                   /product= "mouse erythropoietin"
XX
PN   WO200166149-A2.
XX
PD   13-SEP-2001.
XX
PF   02-MAR-2001; 2001WO-US06953.
XX
PR   03-MAR-2000; 2000US-187236P.
PR   16-JAN-2001; 2001US-261751P.
XX
PA   (VALE-) VALENTIS INC.
XX
PI   Fewell JG,  Maclaughlin F,  Smith LC,  Nicol F,  Rolland A;
XX
DR   WPI; 2001-638995/73.
XX
PT   Nucleic acid formulation for gene delivery to a muscle or tumour tissue
PT   to treat cancer, or infectious disease in a mammal, comprises a nucleic
PT   acid and non-encapsulating anionic polymer such as poly-L-glutamate -
XX
PS   Example 4; Page 90-91; 98pp; English.
XX
CC   The present sequence is that of expression plasmid pEP1403, which
CC   encodes mouse erythropoietin (EPO).  The IFN-alpha coding sequence
CC   was inserted into the Valentis plasmid backbone containing a 107 bp
CC   5' untranslated region, a 117 bp synthetic intron, the human growth
CC   hormone polyadenylation signal, a pUC12 origin of replication and a
CC   kanamycin resistance gene, such that the EPO gene was driven by the
CC   cytomegalovirus enhancer/promoter.  The resulting plasmid, pEP1403,
CC   was formulated with poly-L-glutamate to produce a gene delivery
CC   vehicle, which was intramuscularly injected into mice, under
CC   electroporation conditions.  Delivery of the plasmid DNA in the
CC   poly-glutamate formulation resulted in considerably higher levels
CC   of expressed protein than when the plasmid DNA was delivered in
CC   saline.  This is an example of a method designed for non-viral
CC   plasmid-based gene therapy.  In this method, a nucleic acid is
CC   formulated with a non-encapsulating anionic polymer, such as
CC   (biodegradable) poly-L-glutamate, which enhances transfection of
CC   the nucleic acid into muscle or tumour tissues, with or without
CC   electroporation, and which also stabilises the nucleic acid during
CC   storage.  The formulations allow for vaccination and treatment of
CC   muscle disorders and serum protein deficiencies, as well as cancer
CC   and infections.  In the case of EPO gene delivery, it may be used
CC   to treat anaemia.
XX
SQ   Sequence 3609 BP; 811 A; 1031 C; 953 G; 814 T; 0 other;

  Query Match           100.0%;  Score 21;  DB 22;  Length 3609;
```

No evidence of plasmid publication prior to 1996

Looks, based on 98/17689, like Valentis is using Nordstrom intron

```
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TTCTTTTTTTCTCTTCACAGG 21
           |||||||||||||||||||||
Db     767 ttctttttttctcttcacagg 787


RESULT   7
AAI70087
ID   AAI70087 standard; DNA; 4276 BP.
XX
AC   AAI70087;
XX
DT   21-DEC-2001  (first entry)
XX
DE   Codon optimised plasmid pFN1645 for human coagulation Factor IX.
XX
KW   Plasmid pFN1645; Factor IX; coagulation; blood clotting; human;
KW   gene delivery; haemophilia B; gene therapy; vaccine; codon usage;
KW   ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             786..2171
FT                   /*tag= a
FT                   /product= "human Factor IX"
XX
PN   WO200166149-A2.
XX
PD   13-SEP-2001.
XX
PF   02-MAR-2001; 2001WO-US06953.
XX
PR   03-MAR-2000; 2000US-187236P.
PR   16-JAN-2001; 2001US-261751P.
XX
PA   (VALE-) VALENTIS INC.
XX
PI   Fewell JG,  Maclaughlin F,  Smith LC,  Nicol F,  Rolland A;
XX
DR   WPI; 2001-638995/73.
XX
PT   Nucleic acid formulation for gene delivery to a muscle or tumour tissue
PT   to treat cancer, or infectious disease in a mammal, comprises a nucleic
PT   acid and non-encapsulating anionic polymer such as poly-L-glutamate -
XX
PS   Claim 77; Page 94-96; 98pp; English.
XX
CC   The present sequence is that of expression plasmid pFN1645, which
CC   encodes human Factor IX (FIX).  The plasmid has a codon optimised
CC   sequence which may be preferred to the natural FIX nucleic acid
CC   sequence (in plasmid pFN0945, see AAI70086) when higher expression
CC   is desired.  In experiments with pFN0945, plasmid DNA was formulated
CC   with poly-L-glutamate to produce a gene delivery vehicle, which was
CC   intramuscularly injected into C57BL/6 mice tibialis, augmented by
CC   electroporation.  The highest expression of human FIX achieved
CC   using this method 280 ng/ml, compared with levels of 160 ng/ml
CC   obtained with naked DNA treatment.  Expression was dose
CC   dependent, and the plasmid was stable and transcriptionally active
CC   in muscle for a prolonged period of time.  Applicability to large
CC   animals (dog) was demonstrated.  Some muscle damage was observed
CC   1 mth after treatment.  This is an example of the method of the
CC   invention for non-viral plasmid-based gene therapy.  In this method,
CC   a nucleic acid is formulated with a non-encapsulating anionic
CC   polymer, such as (biodegradable) poly-L-glutamate, which not only
CC   enhances transfection of the nucleic acid into muscle or tumour
CC   tissues, with or without electroporation, but also stabilises the
CC   nucleic acid during storage.  The formulations allow for
CC   vaccination and treatment of muscle disorders and serum protein
CC   deficiencies, as well as cancer and infections.  In the case of
CC   FIX gene delivery, it may be used to treat haemophilia B.
```

```
XX
SQ   Sequence 4276 BP; 1059 A; 1092 C; 1120 G; 1005 T; 0 other;


  Query Match             100.0%;  Score 21;  DB 22;  Length 4276;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TTCTTTTTTTCTCTTCACAGG 21
           |||||||||||||||||||||
Db     749 ttctttttttctcttcacagg 769


RESULT   8
AAI70086
ID   AAI70086 standard; DNA; 4496 BP.
XX
AC   AAI70086;
XX
DT   21-DEC-2001  (first entry)
XX
DE   Plasmid pFN0945 encoding human coagulation Factor IX.
XX
KW   Plasmid pFN0945; Factor IX; coagulation; blood clotting; human;
KW   gene delivery; haemophilia B; gene therapy; vaccine; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             782..2167
FT                   /*tag=  a
FT                   /product= "human Factor IX"
XX
PN   WO200166149-A2.
XX
PD   13-SEP-2001.
XX
PF   02-MAR-2001; 2001WO-US06953.
XX
PR   03-MAR-2000; 2000US-187236P.
PR   16-JAN-2001; 2001US-261751P.
XX
PA   (VALE-) VALENTIS INC.
XX
PI   Fewell JG,  Maclaughlin F,  Smith LC,  Nicol F,  Rolland A;
XX
DR   WPI; 2001-638995/73.
XX
PT   Nucleic acid formulation for gene delivery to a muscle or tumour tissue
PT   to treat cancer, or infectious disease in a mammal, comprises a nucleic
PT   acid and non-encapsulating anionic polymer such as poly-L-glutamate   -
XX
PS   Claim 77; Page 92-93; 98pp; English.
XX
CC   The present sequence is that of expression plasmid pFN0945, which
CC   encodes human Factor IX (FIX).  The plasmid was formulated with
CC   poly-L-glutamate to produce a gene delivery vehicle, which was
CC   intramuscularly injected into C57BL/6 mice tibialis, augmented by
CC   electroporation.  The highest expression of human FIX achieved
CC   using this method 280 ng/ml, compared with levels of 160 ng/ml
CC   obtained with naked DNA treatment.  Expression was dose
CC   dependent, and the plasmid was stable and transcriptionally active
CC   in muscle for a prolonged period of time.  Applicability to large
CC   animals (dog) was demonstrated.  Some muscle damage was observed
CC   1 mth after treatment.  This is an example of the method of the
CC   invention for non-viral plasmid-based gene therapy.  In this method,
CC   a nucleic acid is formulated with a non-encapsulating anionic
CC   polymer, such as (biodegradable) poly-L-glutamate, which not only
CC   enhances transfection of the nucleic acid into muscle or tumour
CC   tissues, with or without electroporation, but also stabilises the
CC   nucleic acid during storage.  The formulations allow for
CC   vaccination and treatment of muscle disorders and serum protein
CC   deficiencies, as well as cancer and infections.  In the case of
CC   FIX gene delivery, it may be used to treat haemophilia B.
XX
SQ   Sequence 4496 BP; 1127 A; 1119 C; 1147 G; 1103 T; 0 other;


  Query Match             100.0%;  Score 21;  DB 22;  Length 4496;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TTCTTTTTTTCTCTTCACAGG 21
           |||||||||||||||||||||
Db     748 ttctttttttctcttcacagg 768


RESULT   9
AAZ40415
ID   AAZ40415 standard; DNA; 5686 BP.
XX
AC   AAZ40415;
XX
DT   15-FEB-2000  (first entry)
XX
DE   Plasmid pIN1143 containing human IL-12 sequence.
XX
KW   Wild type; human; interferon-alpha; plasmid; cytomegalovirus; CMV;
KW   promoter; growth hormone; untranslated region; UTR; mammal; disease;
KW   cancer; intron; ss.
XX
OS   Synthetic.
XX
PN   WO9947678-A2.
XX
PD   23-SEP-1999.
XX
PF   12-MAR-1999;   99WO-US05394.
XX
PR   19-MAR-1998;   98US-0078654.
XX
PA   (GENE-) GENEMEDICINE INC.
XX
PI   Nordstrom J,  Pericle F,  Rolland A,  Ralston R;
XX
DR   WPI; 1999-562116/47.
XX
PT   New plasmids containing an interferon-alpha coding sequence, used for
PT   the treatment of a mammalian condition or disease, particularly cancer
PT   -
XX
PS   Disclosure; Fig 2; 137pp; English.
XX
CC   The invention relates to a novel plasmid comprising a cytomegalovirus
CC   (CMV) promoter transcriptionally linked with an interferon alpha
CC   (IFN-alpha) coding sequence, and a growth hormone 3'-untranslated
CC   region (UTR).  This sequence represents the plasmid pIN1143 which
CC   contains the human interleukin 12 (IL-12) gene. The plasmids can be
CC   used for treating a mammalian condition or disease, particularly cancer.
XX
SQ   Sequence 5686 BP; 1367 A; 1517 C; 1446 G; 1356 T; 0 other;


  Query Match             100.0%;  Score 21;  DB 20;  Length 5686;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TTCTTTTTTTCTCTTCACAGG 21
           |||||||||||||||||||||
Db    1519 ttctttttttctcttcacagg 1539


RESULT  10
AAZ40417
ID   AAZ40417 standard; DNA; 5966 BP.
XX
```

```
AC   AAZ40417;
XX
DT   15-FEB-2000  (first entry)
XX
DE   Plasmid pIN0961 containing mouse IL-12 sequence.
XX
KW   Wild type; human; interferon-alpha; plasmid; cytomegalovirus; CMV;
KW   promoter; growth hormone; untranslated region; UTR; mammal; disease;
KW   cancer; intron; ss.
XX
OS   Synthetic.
XX
PN   WO9947678-A2.
XX
PD   23-SEP-1999.
XX
PF   12-MAR-1999;   99WO-US05394.
XX
PR   19-MAR-1998;   98US-0078654.
XX
PA   (GENE-) GENEMEDICINE INC.
XX
PI   Nordstrom J,  Pericle F,  Rolland A,  Ralston R;
XX
DR   WPI; 1999-562116/47.
XX
PT   New plasmids containing an interferon-alpha coding sequence, used for
PT   the treatment of a mammalian condition or disease, particularly cancer
PT   -
XX
PS   Disclosure; Fig 5; 137pp; English.
XX
CC   The invention relates to a novel plasmid comprising a cytomegalovirus
CC   (CMV) promoter transcriptionally linked with an interferon alpha
CC   (IFN-alpha) coding sequence, and a growth hormone 3'-untranslated
CC   region (UTR).  This sequence represents the plasmid pIN0961 which
CC   contains the mouse interleukin 12 (IL-12) gene. The plasmids can be
CC   used for treating a mammalian condition or disease, particularly cancer.
XX
SQ   Sequence 5966 BP; 1421 A; 1627 C; 1542 G; 1376 T; 0 other;


  Query Match             100.0%;  Score 21;  DB 20;  Length 5966;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TTCTTTTTTTCTCTTCACAGG 21
          |||||||||||||||||||||
Db   2861 ttcttttttctcttcacagg 2881


RESULT  11
AAZ09247/c
ID   AAZ09247 standard; cDNA; 1462 BP.
XX
AC   AAZ09247;
XX
DT   25-OCT-1999  (first entry)
XX
DE   Human CARD-4L partial cDNA.
XX
KW   CARD-3; caspase recruitment domain; CARD-4; regulation; detection;
KW   caspase activation; detection; screening; therapy; diagnosis; disease;
KW   apoptotic cell death; Fas/APO-1 receptor complex; TNF receptor complex;
KW   cancer; follicular lymphoma; carcinoma; p53 mutation; viral infection;
KW   hormone-dependent tumour; autoimmune disorder; Alzheimer's disease;
KW   systemic lupus erythematosis; immune-mediated glomerulonephritis; stroke;
KW   Parkinson's disease; amyotrophic lateral sclerosis; retinitis pigmentosa;
KW   spinal muscular dystrophy; cerebellar degeneration; anaemia; drug;
KW   myelodysplastic syndrome; myocardial infarction; cell proliferation;
KW   cell differentiation; cell survival; CARD-4L; CARD-4S; CARD-4Y;
KW   CARD-4Z; human; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1186
FT                   /*tag=  a
FT                   /codon_start= 2
FT                   /note= "Partial CARD-4L coding sequence"
XX
PN   WO9940102-A1.
XX
PD   12-AUG-1999.
XX
PF   05-FEB-1999;   99WO-US02544.
XX
PR   08-DEC-1998;   98US-0207359.
PR   06-FEB-1998;   98US-0019942.
PR   17-JUN-1998;   98US-0099041.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Bertin J;
XX
DR   WPI; 1999-494269/41.
DR   P-PSDB; AAY31141.
XX
PT   Novel CARD-3 and CARD-4 genes and polypeptides used or treating
PT   regulation of cellular proliferation and differentiation and cell
PT   survival
XX
PS   Example 2; Fig 3; 181pp; English.
XX
CC   This invention describes the isolation of novel human caspase
CC   recruitment domain, CARD-3 and CARD-4 polynucleotides and proteins and a
CC   partial murine CARD-4L protein and genes. The genes and proteins of
CC   the invention are involved in the regulation of caspase activation.
CC   The caspase recruitment domain (CARD) polynucleotides, polypeptides,
CC   homologues and antibodies can be used in screening assays, detection
CC   assays, predictive medicine and therapeutic and prophylactic methods of
CC   treatment. The methods may be used to diagnose and treat patients which
CC   are suffering from a disorder associated with abnormal level or rate of
CC   apoptotic cell death, abnormal activity of the Fas/APO-1 receptor
CC   complex, abnormal activity of the TNF receptor complex, or abnormal
CC   activity of a caspase. Diseases that may be treated include cancer
CC   (particularly follicular lymphoma, carcinomas associated with mutations
CC   in p53 and hormone-dependent tumours), autoimmune disorders (e.g.
CC   systemic lupus erythematosis, immune-mediated glomerulonephritis), viral
CC   infections, Alzheimer's disease, Parkinson's disease, amyotrophic lateral
CC   sclerosis, retinitis pigmentosa, spinal muscular dystrophy, cerebellar
CC   degeneration, anaemia, myelodysplastic syndrome, myocardial infarction,
CC   and stroke. CARD-3 protein interacts with other cellular proteins, and so
CC   can be used for regulation of cellular proliferation and differentiation
CC   and cell survival. The CARD proteins may also be used to for screen drugs
CC   or compounds which modulate their activity. The CARD-4 gene can express a
CC   long transcript that encodes CARD-4L, a short transcript that encodes
CC   CARD-4S or two CARD-4 splice variants, CARD-4Y and CARD-4Z. This sequence
CC   encodes the human CARD-4L protein described in the method of the
CC   invention. This sequence represents the 3'-end fragment of the CARD-4L
CC   coding region, represented in Figure 3, however the specification
CC   describes the full length CARD-4L cDNA sequence which encodes a 953
CC   amino acid protein.
XX
SQ   Sequence 1462 BP; 389 A; 391 C; 389 G; 292 T; 1 other;


  Query Match             87.6%;  Score 18.4;  DB 20;  Length 1462;
  Best Local Similarity   95.0%;  Pred. No. 1.7e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TTCTTTTTTTCTCTTCACAG 20
          || |||||||||||||||||
Db   1451 TTTTTTTTTTCTCTTCACAG 1432
```

```
RESULT  12
AAF30002/c
ID   AAF30002 standard; cDNA; 3382 BP.
XX
AC   AAF30002;
XX
DT   23-APR-2001  (first entry)
XX
DE   Human CARD-4L (long form) cDNA.
XX
KW   CARD-4L; caspase recruitment domain; human; cancer; infection;
KW   autoimmune disease; neurological disease; haematological disease;
KW   immune disease; inflammation; antitumour; antiseptic;
KW   immunomodulator; antiinflammatory; apoptosis; diagnosis;
KW   gene therapy; chromosome 7; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             245..3106
FT                   /*tag= a
FT                   /note= "the open reading frame is also specifically
FT                          claimed in Claim 1(a)"
XX
PN   WO200100826-A2.
XX
PD   04-JAN-2001.
XX
PF   28-JUN-2000; 2000WO-US17691.
XX
PR   28-JUN-1999;   99US-0340620.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Bertin J;
XX
DR   WPI; 2001-061973/07.
DR   P-PSDB; AAB20080.
XX
PT   Isolated intracellular proteins predicted to be involved in regulating
PT   caspase activation are used for diagnosis and treatment of e.g. cancer,
PT   viral infections, autoimmune diseases, neurological diseases and
PT   haematological disorders -
XX
PS   Claim 1(a); Fig 3; 208pp; English.
XX
CC   The present sequence is that of cDNA encoding human caspase
CC   recruitment domain 4 long form (CARD-4L, see AAB20080). The cDNA
CC   was isolated from a human umbilical vein endothelial library using
CC   a partial CARD-4S clone as probe. Plasmid pC4L1 containing CARD-4L
CC   cDNA is deposited as ATCC 203035. The human CARD-4 gene (see
CC   AAF30011) maps to chromosome 7. CARD-4 exists in at least 4 forms,
CC   i.e. the long form CARD-4L, the short form CARD-4S (see AAB20081),
CC   and splice variants CARD-4Y (see AAB20082) and CARD-4Z (see
CC   AAB20082). CARD-4 is an intracellular protein predicted to be
CC   involved in regulating caspase activation. It activates the
CC   NF-kappaB pathway and enhances caspase 9-mediated cell death.
CC   Methods of diagnosing and treating patients suffering from a
CC   disorder associated with an abnormal level or rate of apoptotic
CC   cell death, abnormal activity of the Fas/APO-1 receptor complex,
CC   abnormal activity of the tumour necrosis factor receptor complex
CC   or abnormal activity of a caspase involve administering a compound
CC   that modulates the expression or activity of CARD-3, CARD-4, CARD-5
CC   or CARD-6 e.g. a small molecule, antisense nucleic acid, ribozyme
CC   or polypeptide. Such disorders include cancer, viral infection,
CC   autoimmune disorders, neurological diseases, haematological
CC   disorders, inflammatory disorders and immune disorders. CARD
CC   nucleic acids can be used to express CARD proteins in a host cell
CC   e.g. for gene therapy applications, to detect a genetic lesion and
CC   to modulate CARD activity.
XX
SQ   Sequence 3382 BP; 775 A; 975 C; 933 G; 693 T; 6 other;

  Query Match             87.6%;  Score 18.4;  DB 22;  Length 3382;
  Best Local Similarity   95.0%;  Pred. No. 1.8e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 TTCTTTTTTTCTCTTCACAG 20
           || ||||||||||||||||||
Db    3371 TTTTTTTTTTCTCTTCACAG 3352


RESULT  13
AAI87537/c
ID   AAI87537 standard; cDNA; 366 BP.
XX
AC   AAI87537;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 7597.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04927.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   P-PSDB; AAO07606.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing
PT   diagnosing and treating e.g. leukaemia, inflammation and immune
PT   disorders -
XX
PS   Claim 1; SEQ ID NO 7597; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 366 BP; 210 A; 44 C; 45 G; 66 T; 1 other;

  Query Match             84.8%;  Score 17.8;  DB 22;  Length 366;
  Best Local Similarity   90.5%;  Pred. No. 2.7e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 TTCTTTTTTTCTCTTCACAGG 21
           |||||||||||||||  || ||
```

```
Db    299 TTCTTTTTTTCTCTTGACGGG 279


RESULT  14
AAK87602/c
ID   AAK87602 standard; DNA; 3810 BP.
XX
AC   AAK87602;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:42414.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
```

```
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure; SEQ ID NO 42414; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 3810 BP; 1081 A; 684 C; 703 G; 1321 T; 21 other;


  Query Match             84.8%;  Score 17.8;  DB 22;  Length 3810;
  Best Local Similarity   90.5%;  Pred. No. 3.2e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 TTCTTTTTTTCTCTTCACAGG 21
           || |||||||||  ||||||||
Db    1951 TTTTTTTTTTCTTTTCACAGG 1931


RESULT  15
AAT43682
ID   AAT43682 standard; DNA; 4098 BP.
XX
AC   AAT43682;
XX
DT   03-FEB-1997  (first entry)
XX
DE   Medium chain-specific acyl-(ACP)-thioesterase genomic clone ClTEg1.
XX
KW   acyl-(ACP)-thioesterase; medium-chain length specificity;
KW   oil seed; softener; pesticide; tenside; cosmetic; transgenic plant; ds.
XX
OS   Cuphea lanceolata.
XX
FH   Key             Location/Qualifiers
FT   exon            1787..2294
FT                   /*tag= a
FT                   /number= 2
FT                   /codon_start= 1797..1799
FT   intron          2295..2657
FT                   /*tag= b
FT                   /number= 2
FT   exon            2658..2791
FT                   /*tag= c
FT                   /number= 3
FT   intron          2792..2897
FT                   /*tag= d
FT                   /number= 3
FT   exon            2898..3011
FT                   /*tag= e
FT                   /number= 4
FT   intron          3012..3131
FT                   /*tag= f
FT                   /number= 4
FT   exon            3132..3303
FT                   /*tag= g
FT                   /number= 5
FT   intron          3304..3390
FT                   /*tag= h
FT                   /number= 5
FT   exon            3391..3459
FT                   /*tag= i
FT                   /number= 6
FT   intron          3460..3671
FT                   /*tag= j
FT                   /number= 6
FT   exon            3672..3941
FT                   /*tag= k
FT                   /number= 7
FT                   /note= "stop codon is at 3942..3944"
XX
PN   WO9506740-A.
XX
PD   09-MAR-1995.
XX
PF   02-SEP-1994;   94WO-EP02935.
XX
PR   03-SEP-1993;   93DE-4329828.
XX
PA   (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI   Martini N,  Schell J,  Toepfer R;
XX
DR   WPI; 1995-115455/15.
DR   P-PSDB; AAW06703.
XX
PT   An acyl-(ACP)-thio:esterase DNA of medium-chain specificity -
PT   isolated from Cuphea lanceolata; for plant transformation to
PT   produce C10:0 fatty acids, useful in the prodn of eg cosmetics.
XX
PS   Claim 13; Page -; 40pp; German.
XX
CC   A primer based on amino acids 277-284 of the acyl-(ACP)-thioesterase
CC   from Umbellularia californica was used with a modified oligo-dT
CC   primer with restriction sites for BstBI, BamHI, HindIII and SalI,
CC   in PCR amplification of a specific acyl-(ACP)-thioesterase
CC   hybridisation probe ("PCR42") from a wild-type Umbellularia californica
CC   cDNA library. Three cDNA clones, designated ClTE13, ClTE5 and ClTE12,
CC   each coding for at least part of a thioesterase with medium-chain
CC   specificity (C10:0-specific) were isolated by screening a Cuphea
CC   lanceolata library with probe PCR42. Then, clone ClTE5 was itself
CC   used as a probe to screen a C.lanceolata genomic DNA library and a
CC   total of 23 clones were identified. Four of the genomic clones were
CC   shown to correspond respectively to PCR42 and the three cDNA clones.
CC   The present sequence is that of the genomic clone designated ClTEg1
CC   which corresponds to cDNA clone ClTE12. The binary vector pNBM99-TEg1
CC   (DSM 8477) comprising a fragment of ClTEg1 is specifically claimed.
CC   The DNA sequences will be useful for transforming oil-producing
```

CC plants (e.g. rapeseed, soya, oil palms) to produce C10:0 fatty acids
CC which are starting materials for softeners, pesticides, tensides and
CC cosmetics.
CC N.B. The nucleotide sequences are referred to throughout the
CC specification by their SEQ.ID. numbers but the sequence listing has
CC not been printed in the original patent application.
XX
SQ Sequence 4098 BP; 1103 A; 808 C; 812 G; 1375 T; 0 other;

Query Match 84.8%; Score 17.8; DB 16; Length 4098;
Best Local Similarity 90.5%; Pred. No. 3.3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TTCTTTTTTTCTCTTCACAGG 21
|| |||||||||||| |||||
Db 2638 ttttttttttctcttaacagg 2658

Search completed: July 21, 2002, 09:55:21
Job time: 6382 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2002, 09:47:18 ; Search time 112.48 Seconds
(without alignments)
45.860 Million cell updates/sec

Title: US-09-754-014-10_COPY_25_45
Perfect score: 21
Sequence: 1 TTCTTTTTTTCTCTTCACAGG 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		21	100.0	30	3	US-09-012-366-7	Sequence 7, Appli
2	c	18.4	87.6	3382	4	US-09-099-041A-7	Sequence 7, Appli
3		17.8	84.8	4098	2	US-08-605-106-4	Sequence 4, Appli
4	c	17.8	84.8	14753	4	US-09-821-736-3	Sequence 3, Appli
5	c	17.4	82.9	1930	2	US-08-455-968E-4	Sequence 4, Appli
6	c	17.4	82.9	2033	2	US-08-455-968E-9	Sequence 9, Appli
7	c	17.4	82.9	5599	2	US-08-477-451-9	Sequence 9, Appli
8		17.4	82.9	5599	2	US-08-477-451-13	Sequence 13, Appl
9	c	17.4	82.9	19932	2	US-08-477-451-25	Sequence 25, Appl
10	c	17	81.0	1225	1	US-08-181-629A-3	Sequence 3, Appli
11	c	17	81.0	5496	1	US-08-181-629A-2	Sequence 2, Appli
12		16.2	77.1	491	4	US-09-020-956-52	Sequence 52, Appl
13		16.2	77.1	491	4	US-09-030-607-52	Sequence 52, Appl
14		16.2	77.1	491	4	US-09-439-313-52	Sequence 52, Appl
15	c	16.2	77.1	508	2	US-08-327-451E-23	Sequence 23, Appl
16	c	16.2	77.1	508	2	US-08-458-109-23	Sequence 23, Appl
17	c	16.2	77.1	910	4	US-09-191-608-2	Sequence 2, Appli
18	c	16.2	77.1	953	1	US-08-197-793-1	Sequence 1, Appli
19	c	16.2	77.1	953	2	US-08-636-176-1	Sequence 1, Appli
20	c	16.2	77.1	953	5	PCT-US95-01618-1	Sequence 1, Appli
21	c	16.2	77.1	1083	4	US-09-247-373B-35	Sequence 35, Appl
22		16.2	77.1	1543	2	US-08-991-946A-4	Sequence 4, Appli
23	c	16.2	77.1	2906	1	US-08-554-612C-49	Sequence 49, Appl
24		16.2	77.1	5467	2	US-08-605-106-7	Sequence 7, Appli
25	c	16.2	77.1	7070	1	US-08-619-554-3	Sequence 3, Appli
26		16.2	77.1	35100	1	US-08-306-691B-19	Sequence 19, Appl
27		16.2	77.1	35100	5	PCT-US93-06251-19	Sequence 19, Appl
28	c	16	76.2	4092	4	US-09-306-595C-5	Sequence 5, Appli
29	c	15.8	75.2	181	2	US-08-256-790-5	Sequence 5, Appli
30	c	15.8	75.2	515	4	US-09-439-313-472	Sequence 472, App
31	c	15.8	75.2	767	1	US-07-697-275-1	Sequence 1, Appli
32		15.8	75.2	1750	4	US-09-262-856A-7	Sequence 7, Appli
33	c	15.8	75.2	2550	2	US-08-884-072-2	Sequence 2, Appli
34	c	15.8	75.2	2550	4	US-09-212-168-2	Sequence 2, Appli
35		15.8	75.2	2816	4	US-09-171-337A-1	Sequence 1, Appli
36	c	15.8	75.2	4527	2	US-08-944-449-8	Sequence 8, Appli
37		15.8	75.2	5935	4	US-09-178-973B-17	Sequence 17, Appl
38		15.8	75.2	5935	4	US-09-354-243B-29	Sequence 29, Appl
39	c	15.4	73.3	1533	1	US-07-721-761A-32	Sequence 32, Appl
40	c	15.4	73.3	1533	1	US-07-978-687-32	Sequence 32, Appl
41	c	15.4	73.3	1533	1	US-08-471-791-12	Sequence 12, Appl
42	c	15.4	73.3	1533	5	PCT-US91-01746-12	Sequence 12, Appl
43	c	15.4	73.3	1533	5	PCT-US91-05801-32	Sequence 32, Appl
44	c	15.4	73.3	1675	1	US-07-688-352C-29	Sequence 29, Appl
45	c	15.4	73.3	1675	2	US-08-474-379C-29	Sequence 29, Appl

ALIGNMENTS

```
RESULT   1
US-09-012-366-7
; Sequence 7, Application US/09012366
; Patent No. 6034072
;  GENERAL INFORMATION:
;    APPLICANT:  Robert Ralston
;    APPLICANT:  Susanne Muller
;    APPLICANT:  Russ Mumper
;    APPLICANT:  William Munger
;    APPLICANT:  Maria Bruno
;    TITLE OF INVENTION:  IL-2 GENE EXPRESSION AND
;    TITLE OF INVENTION:  DELIVERY SYSTEMS AND USES
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  FastSEQ for Windows 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/012,366
;      FILING DATE:  January 23, 1998
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/039,709
;      FILING DATE:  February 10, 1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkman, Charles S.
;      REGISTRATION NUMBER:  38,077
;      REFERENCE/DOCKET NUMBER:  230/214
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  30 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-012-366-7
```

```
  Query Match             100.0%;  Score 21;  DB 3;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 1.7;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TTCTTTTTTTCTCTTCACAGG 21
           |||||||||||||||||||||
Db      10 TTCTTTTTTTCTCTTCACAGG 30


RESULT   2
US-09-099-041A-7/c
; Sequence 7, Application US/09099041A
; Patent No. 6340576
; GENERAL INFORMATION:
;  APPLICANT: Bertin, John
;  TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
;  TITLE OF INVENTION:  PROTEIN FAMILY AND USES THEREOF
;  FILE REFERENCE: 07334-076001
;  CURRENT APPLICATION NUMBER: US/09/099,041A
;  CURRENT FILING DATE:  1998-06-17
;  PRIOR APPLICATION NUMBER: 09/019,942
;  PRIOR FILING DATE: 1998-02-06
;  NUMBER OF SEQ ID NOS: 37
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
;   LENGTH: 3382
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (245)...(3103)
US-09-099-041A-7


  Query Match             87.6%;  Score 18.4;  DB 4;  Length 3382;
  Best Local Similarity   95.0%;  Pred. No. 26;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 TTCTTTTTTTCTCTTCACAG 20
           || ||||||||||||||||||
Db    3371 TTTTTTTTTTCTCTTCACAG 3352


RESULT   3
US-08-605-106-4
; Sequence 4, Application US/08605106
; Patent No. 5910631
;  GENERAL INFORMATION:
;    APPLICANT:  Topfer, R.
;    APPLICANT:  Martini, N.
;    APPLICANT:  Schell, J.
;    TITLE OF INVENTION:  MEDIUM CHAIN-SPECIFIC THIOESTERS
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schwegman, Lundberg, Woessner & Kluth, P.A.
;      STREET:  P.O. Box 2938
;      CITY:  Minneapolis
;      STATE:  MN
;      COUNTRY:  USA
;      ZIP:  55402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/605,106
;      FILING DATE:  23-SEPT-1996
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP94/02935
;      FILING DATE:  01-MAR-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Woessner, Warren D
;      REGISTRATION NUMBER:  30,440
;      REFERENCE/DOCKET NUMBER:  235.001US1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  612-373-6900
;      TELEFAX:  612-339-3061
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4098 Base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  : DNS (genomic)
;    HYPOTHETICAL:  No
;    ANTI-SENSE:  No
;    ORIGINAL SOURCE:
;      ORGANISM:  Cuphea lanceolata
;    IMMEDIATE SOURCE:
;      LIBRARY:  genomic Lambda FIX II
;      CLONE:  ClTEg1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(1797..2294, 2658..2791, 2898..3011, 3132
;      LOCATION:  ..3303, 3391..3459, 3672..3941)
;    FEATURE:
;      NAME/KEY:  Startcodon
;      LOCATION:  1797..1799
;    FEATURE:
;      NAME/KEY:  exon II
;      LOCATION:  1787..2294
;    FEATURE:
;      NAME/KEY:  intron II
;      LOCATION:  2295..2657
;    FEATURE:
;      NAME/KEY:  exon III
;      LOCATION:  2658..2791
;    FEATURE:
;      NAME/KEY:  intron III
;      LOCATION:  2792..2897
;    FEATURE:
;      NAME/KEY:  exon IV
;      LOCATION:  2898..3011
;    FEATURE:
;      NAME/KEY:  intron IV
;      LOCATION:  3012..3131
;    FEATURE:
;      NAME/KEY:  exon V
;      LOCATION:  3132..3303
;    FEATURE:
;      NAME/KEY:  intron V
;      LOCATION:  3304..3390
;    FEATURE:
;      NAME/KEY:  exon VI
;      LOCATION:  3391..3459
;    FEATURE:
;      NAME/KEY:  intron VI
;      LOCATION:  3460..3671
;    FEATURE:
;      NAME/KEY:  exon VII
;      LOCATION:  3672..3941
;    FEATURE:
;      NAME/KEY:  Stopcodon
;      LOCATION:  3942..3944
US-08-605-106-4


  Query Match             84.8%;  Score 17.8;  DB 2;  Length 4098;
  Best Local Similarity   90.5%;  Pred. No. 46;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 TTCTTTTTTTCTCTTCACAGG 21
```

```
          || |||||||||||| |||||
Db   2638 TTTTTTTTTTCTCTTAACAGG 2658


RESULT   4
US-09-821-736-3/c
; Sequence 3, Application US/09821736
; Patent No. 6326182
; GENERAL INFORMATION:
;  APPLICANT: WEBSTER, Marion et al
;  TITLE OF INVENTION: ISOLATED HUMAN LIPASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN LIPASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001216
;  CURRENT APPLICATION NUMBER: US/09/821,736
;  CURRENT FILING DATE:  2001-03-30
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 14753
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(14753)
;   OTHER INFORMATION: n = A,T,C or G
US-09-821-736-3


  Query Match             84.8%;  Score 17.8;  DB 4;  Length 14753;
  Best Local Similarity   90.5%;  Pred. No. 50;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 TTCTTTTTTTCTCTTCACAGG 21
           |||||||| ||||||||| ||
Db   4890 TTCTTTTTCTCTCTTCACTGG 4870


RESULT   5
US-08-455-968E-4/c
; Sequence 4, Application US/08455968E
; Patent No. 5874283
;  GENERAL INFORMATION:
;    APPLICANT:  Harrington, John L.
;    APPLICANT:  Hsieh, Chih-Lin
;    APPLICANT:  Lieber, Michael
;    TITLE OF INVENTION:  Mammalian Flap-Specific Endonuclease
;    NUMBER OF SEQUENCES:  63
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/455,968E
;      FILING DATE:  30-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Apple, Randolph T.
;      REGISTRATION NUMBER:  36,429
;      REFERENCE/DOCKET NUMBER:  18985-000100
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-576-0200
;      TELEFAX:  415-576-0300
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1930 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-455-968E-4


  Query Match             82.9%;  Score 17.4;  DB 2;  Length 1930;
  Best Local Similarity   94.7%;  Pred. No. 62;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 TTCTTTTTTTCTCTTCACA 19
           || ||||||||||||||||
Db   1914 TTTTTTTTTTCTCTTCACA 1896


RESULT   6
US-08-455-968E-9/c
; Sequence 9, Application US/08455968E
; Patent No. 5874283
;  GENERAL INFORMATION:
;    APPLICANT:  Harrington, John L.
;    APPLICANT:  Hsieh, Chih-Lin
;    APPLICANT:  Lieber, Michael
;    TITLE OF INVENTION:  Mammalian Flap-Specific Endonuclease
;    NUMBER OF SEQUENCES:  63
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/455,968E
;      FILING DATE:  30-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Apple, Randolph T.
;      REGISTRATION NUMBER:  36,429
;      REFERENCE/DOCKET NUMBER:  18985-000100
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-576-0200
;      TELEFAX:  415-576-0300
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2033 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  104..1237
US-08-455-968E-9


  Query Match             82.9%;  Score 17.4;  DB 2;  Length 2033;
  Best Local Similarity   94.7%;  Pred. No. 63;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 TTCTTTTTTTCTCTTCACA 19
           || ||||||||||||||||
Db   2017 TTTTTTTTTTCTCTTCACA 1999


RESULT   7
```

```
US-08-477-451-9/c
; Sequence 9, Application US/08477451
; Patent No. 5928865
;  GENERAL INFORMATION:
;    APPLICANT:  Covacci, Antonello
;    TITLE OF INVENTION:  Helicobacter Pylori Cagi Region
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Chiron Corporation
;      STREET:  4560 Horton Street
;      CITY:  Emeryville
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94608-2916
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/477,451
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McClung, Barbara G.
;      REGISTRATION NUMBER:  33,113
;      REFERENCE/DOCKET NUMBER:  0335.002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  510-601-2708
;      TELEFAX:  510-655-3542
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5599 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-477-451-9


  Query Match             82.9%;  Score 17.4;  DB 2;  Length 5599;
  Best Local Similarity   94.7%;  Pred. No. 68;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 TTCTTTTTTTCTCTTCACA 19
           |||||||||||||||| ||
Db    1695 TTCTTTTTTTCTCTTCTCA 1677


RESULT   8
US-08-477-451-13
; Sequence 13, Application US/08477451
; Patent No. 5928865
;  GENERAL INFORMATION:
;    APPLICANT:  Covacci, Antonello
;    TITLE OF INVENTION:  Helicobacter Pylori Cagi Region
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Chiron Corporation
;      STREET:  4560 Horton Street
;      CITY:  Emeryville
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94608-2916
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/477,451
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McClung, Barbara G.
;      REGISTRATION NUMBER:  33,113
;      REFERENCE/DOCKET NUMBER:  0335.002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  510-601-2708
;      TELEFAX:  510-655-3542
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5599 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-477-451-13


  Query Match             82.9%;  Score 17.4;  DB 2;  Length 5599;
  Best Local Similarity   94.7%;  Pred. No. 68;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 TTCTTTTTTTCTCTTCACA 19
           |||||||||||||||| ||
Db    3905 TTCTTTTTTTCTCTTCTCA 3923


RESULT   9
US-08-477-451-25/c
; Sequence 25, Application US/08477451
; Patent No. 5928865
;  GENERAL INFORMATION:
;    APPLICANT:  Covacci, Antonello
;    TITLE OF INVENTION:  Helicobacter Pylori Cagi Region
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Chiron Corporation
;      STREET:  4560 Horton Street
;      CITY:  Emeryville
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94608-2916
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/477,451
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McClung, Barbara G.
;      REGISTRATION NUMBER:  33,113
;      REFERENCE/DOCKET NUMBER:  0335.002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  510-601-2708
;      TELEFAX:  510-655-3542
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  19932 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-477-451-25


  Query Match             82.9%;  Score 17.4;  DB 2;  Length 19932;
  Best Local Similarity   94.7%;  Pred. No. 74;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 TTCTTTTTTTCTCTTCACA 19
           |||||||||||||||| ||
```

```
Db   5252 TTCTTTTTTTCTCTTCTCA 5234


RESULT  10
US-08-181-629A-3/c
; Sequence 3, Application US/08181629A
; Patent No. 5472872
;  GENERAL INFORMATION:
;    APPLICANT:  Swaminathan, Neela
;    APPLICANT:  Van Etten, James
;    APPLICANT:  Mead, David
;    APPLICANT:  Skowron, Piotr
;    TITLE OF INVENTION:  Recombinant CviJI Restriction Endonuclease
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/181,629A
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  31504
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1225 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(1..33, 55..1128)
US-08-181-629A-3


  Query Match             81.0%;  Score 17;  DB 1;  Length 1225;
  Best Local Similarity   100.0%;  Pred. No. 87;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TTCTTTTTTTCTCTTCA 17
          |||||||||||||||||
Db    139 TTCTTTTTTTCTCTTCA 123


RESULT  11
US-08-181-629A-2/c
; Sequence 2, Application US/08181629A
; Patent No. 5472872
;  GENERAL INFORMATION:
;    APPLICANT:  Swaminathan, Neela
;    APPLICANT:  Van Etten, James
;    APPLICANT:  Mead, David
;    APPLICANT:  Skowron, Piotr
;    TITLE OF INVENTION:  Recombinant CviJI Restriction Endonuclease
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/181,629A
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  31504
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5496 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-181-629A-2


  Query Match             81.0%;  Score 17;  DB 1;  Length 5496;
  Best Local Similarity   100.0%;  Pred. No. 97;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TTCTTTTTTTCTCTTCA 17
          |||||||||||||||||
Db   1207 TTCTTTTTTTCTCTTCA 1191


RESULT  12
US-09-020-956-52
; Sequence 52, Application US/09020956
; Patent No. 6261562
;  GENERAL INFORMATION:
;    APPLICANT:  Xu, Jiangchun
;    APPLICANT:  Dillin, Davin C.
;    TITLE OF INVENTION:  COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
;    NUMBER OF SEQUENCES:  178
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/020,956
;      FILING DATE:  09-FEB-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.427C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
```

```
;  INFORMATION FOR SEQ ID NO:  52:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  491 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-09-020-956-52


  Query Match             77.1%;  Score 16.2;  DB 4;  Length 491;
  Best Local Similarity   85.7%;  Pred. No. 1.7e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 TTCTTTTTTTCTCTTCACAGG 21
          ||||||||||  | || |||||
Db    387 TTCTTTTTTTTTTTTTACAGG 407


RESULT  13
US-09-030-607-52
; Sequence 52, Application US/09030607
; Patent No. 6262245
;  GENERAL INFORMATION:
;    APPLICANT:  Xu, Jiangchun
;    APPLICANT:  Dillon, Davin C.
;    TITLE OF INVENTION:  COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO
;    NUMBER OF SEQUENCES:  224
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/030,607
;      FILING DATE:  25-FEB-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.427C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  52:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  491 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-09-030-607-52


  Query Match             77.1%;  Score 16.2;  DB 4;  Length 491;
  Best Local Similarity   85.7%;  Pred. No. 1.7e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 TTCTTTTTTTCTCTTCACAGG 21
          ||||||||||  | || |||||
Db    387 TTCTTTTTTTTTTTTTACAGG 407


RESULT  14
US-09-439-313-52
; Sequence 52, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Harlocker, Susan Louise
;  APPLICANT:  Jiang Yuqui
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Kalos, Michael
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Retter, Mark
;  APPLICANT:  Solk, John
;  APPLICANT:  Day, Craig
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE: 210121.427C9
;  CURRENT APPLICATION NUMBER: US/09/439,313
;  CURRENT FILING DATE:  1999-11-12
;  NUMBER OF SEQ ID NOS: 575
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 52
;   LENGTH: 491
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(491)
;   OTHER INFORMATION: n = A,T,C or G
US-09-439-313-52


  Query Match             77.1%;  Score 16.2;  DB 4;  Length 491;
  Best Local Similarity   85.7%;  Pred. No. 1.7e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 TTCTTTTTTTCTCTTCACAGG 21
          ||||||||||  | || |||||
Db    387 ttctttttttttttttacagg 407


RESULT  15
US-08-327-451E-23/c
; Sequence 23, Application US/08327451E
; Patent No. 5910630
;  GENERAL INFORMATION:
;    APPLICANT:  Davies, Maelor
;    APPLICANT:  Hawkins, Deborah
;    APPLICANT:  Nelsen, Janet
;    APPLICANT:  Lassner, Michael
;    TITLE OF INVENTION:  PLANT LYSOPHOSPHATIDIC
;    TITLE OF INVENTION:  ACID ACYLTRANSFERASES
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Calgene, Inc.
;      STREET:  1920 Fifth Street
;      CITY:  Davis
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  95616
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 MB
;      COMPUTER:  IBM PC
;      OPERATING SYSTEM:  Windows NT 4.0
;      SOFTWARE:  Microsoft Word For Windows 7.0a
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/327,451E
;      FILING DATE:  21-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/254,404
```

```
;      FILING DATE:  06-JUN-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/231,196
;      FILING DATE:  21-APR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/224,625
;      FILING DATE:  06-APR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carl J. Schwedler
;      REGISTRATION NUMBER:  36,924
;      REFERENCE/DOCKET NUMBER:  CGNE 106-3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (530) 753-6313
;      TELEFAX:  (530) 753-1510
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  508 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
US-08-327-451E-23


  Query Match             77.1%;  Score 16.2;  DB 2;  Length 508;
  Best Local Similarity   85.7%;  Pred. No. 1.7e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 TTCTTTTTTTCTCTTCACAGG 21
           |||||||||| ||||||  ||
Db     496 TTCTTTTTTTTTCTTCAACGG 476


Search completed: July 21, 2002, 09:47:21
Job time: 11952 sec
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2002, 09:11:05 ; Search time 3274.61 Seconds
(without alignments)
86.556 Million cell updates/sec

Title: US-09-754-014-10_COPY_25_45
Perfect score: 21
Sequence: 1 TTCTTTTTTTCTCTTCACAGG 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	19.4	92.4	537	10	BI219527	BI219527 602936604
c	2	19.4	92.4	691	10	BG499610	BG499610 602546774
	3	18.4	87.6	242	9	BB015899	BB015899 BB015899
	4	18.4	87.6	301	9	BB540713	BB540713 BB540713
	5	18.4	87.6	327	9	BB125849	BB125849 BB125849
c	6	18.4	87.6	335	9	AA063675	AA063675 T3357 MVA
	7	18.4	87.6	341	12	AQ908298	AQ908298 GSSTc0505
	8	18.4	87.6	349	12	BH057848	BH057848 RPCI-24-3
	9	18.4	87.6	407	9	AW493746	AW493746 UI-M-BH3-
c	10	18.4	87.6	511	10	BG140084	BG140084 EST480526
c	11	18.4	87.6	522	9	AA677704	AA677704 zj72h09.s
	12	18.4	87.6	561	10	BM217071	BM217071 C0890E05-
c	13	18.4	87.6	577	10	BM030926	BM030926 495476 MA
c	14	18.4	87.6	593	9	AI671885	AI671885 wb41b12.x
c	15	18.4	87.6	679	12	AG055132	AG055132 Pan trogl
c	16	18.4	87.6	855	12	CNS05FV8	AL335501 Tetraodon
	17	18	85.7	289	9	BB365025	BB365025 BB365025
	18	18	85.7	456	12	B35225	B35225 HS-1027-A2-
c	19	18	85.7	520	10	BG659544	BG659544 TgESTzya2
c	20	18	85.7	652	10	BG709019	BG709019 602675169
c	21	18	85.7	907	12	CNS04AC9	AL281682 Tetraodon
	22	18	85.7	931	9	BE039837	BE039837 OC08G12 O
	23	18	85.7	1014	9	AL540909	AL540909 AL540909
	24	17.8	84.8	193	10	BI424873	BI424873 sah49h02.
c	25	17.8	84.8	243	10	BF023701	BF023701 EtESTea52
	26	17.8	84.8	271	9	AV079879	AV079879 AV079879
	27	17.8	84.8	276	10	BE993065	BE993065 UI-M-BZ1-
c	28	17.8	84.8	332	10	BM446523	BM446523 1IL8H11.a
c	29	17.8	84.8	340	9	BB104566	BB104566 BB104566
c	30	17.8	84.8	347	10	BI402376	BI402376 MI-P-CP0-
c	31	17.8	84.8	353	12	BH289808	BH289808 CH230-2O2
	32	17.8	84.8	357	9	BE106435	BE106435 UI-R-BO1-
	33	17.8	84.8	372	9	AI136137	AI136137 UI-R-C2p-
c	34	17.8	84.8	393	10	BG541755	BG541755 602569639
	35	17.8	84.8	409	9	AW785838	AW785838 117303 MA
	36	17.8	84.8	419	9	AW514225	AW514225 hd24a09.x
	37	17.8	84.8	453	12	AQ077923	AQ077923 CIT-HSP-2
c	38	17.8	84.8	458	12	AQ803358	AQ803358 HS_3139_A
c	39	17.8	84.8	469	10	BF532479	BF532479 602074528
	40	17.8	84.8	477	12	AZ102514	AZ102514 RPCI-23-2
	41	17.8	84.8	477	12	BH340152	BH340152 CH230-41E
	42	17.8	84.8	479	12	AQ208834	AQ208834 HS_3056_B
c	43	17.8	84.8	507	12	AZ714340	AZ714340 RPCI-24-1
c	44	17.8	84.8	531	12	AZ868125	AZ868125 2M0179B17
	45	17.8	84.8	534	12	AZ144116	AZ144116 SP_0042_B

ALIGNMENTS

RESULT 1
BI219527/c
LOCUS BI219527 537 bp mRNA linear EST 11-JUL-2001
DEFINITION 602936604F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:5099929 5', mRNA sequence.
ACCESSION BI219527
VERSION BI219527.1 GI:14672971
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 537)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11240 row: g column: 02
High quality sequence stop: 485.
FEATURES Location/Qualifiers
source 1. .537
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5099929"
/clone_lib="NCI_CGAP_Li9"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.9 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 108 a 145 c 161 g 123 t

```
ORIGIN


  Query Match             92.4%;  Score 19.4;  DB 10;  Length 537;
  Best Local Similarity   95.2%;  Pred. No. 1.9e+04;
  Matches   20;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TTCTTTTTTTCTCTTCACAGG 21
          ||||||||||||||| |||||
Db    534 TTCTTTTTTTCTCTTAACAGG 514


RESULT   2
BG499610/c
LOCUS       BG499610                691 bp    mRNA    linear   EST 27-MAR-2001
DEFINITION  602546774F1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4669003 5',
            mRNA sequence.
ACCESSION   BG499610
VERSION     BG499610.1  GI:13461127
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 691)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1480  row: c  column: 20
            High quality sequence stop: 211.
FEATURES             Location/Qualifiers
     source          1. .691
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4669003"
                     /clone_lib="NIH_MGC_60"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: prostate; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); Double-stranded cDNA was prepared from cell line RNA.
                     5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).  Average
                     insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies
                     contained inserts  by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
                     Library."
BASE COUNT      304 a    103 c    220 g     61 t      3 others
ORIGIN


  Query Match             92.4%;  Score 19.4;  DB 10;  Length 691;
  Best Local Similarity   95.2%;  Pred. No. 1.7e+04;
  Matches   20;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TTCTTTTTTTCTCTTCACAGG 21
          ||||||||||||||| ||||
Db    256 TTCTTTTTTTCTCTTCTCAGG 236


RESULT   3
BB015899
LOCUS       BB015899                242 bp    mRNA    linear   EST 22-JUN-2000
DEFINITION  BB015899 RIKEN full-length enriched, adult male testis (DH10B) Mus
            musculus cDNA clone 4930556M05 3', mRNA sequence.
ACCESSION   BB015899
VERSION     BB015899.1  GI:8186874
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 242)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .242
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="4930556M05"
                     /clone_lib="RIKEN full-length enriched, adult male testis
                     (DH10B)"
                     /sex="male"
                     /tissue_type="testis"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
```

primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."
BASE COUNT 78 a 45 c 33 g 86 t
ORIGIN

Query Match 87.6%; Score 18.4; DB 9; Length 242;
Best Local Similarity 95.0%; Pred. No. 4.8e+04;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TTCTTTTTTTCTCTTCACAG 20
||||||||||||| |||||||
Db 96 TTCTTTTTTTCTATTCACAG 115

RESULT 4
BB540713
LOCUS BB540713 301 bp mRNA linear EST 31-JUL-2000
DEFINITION BB540713 RIKEN full-length enriched, 0 day neonate eyeball Mus musculus cDNA clone E130107M24 3', mRNA sequence.
ACCESSION BB540713
VERSION BB540713.1 GI:9611936
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 301)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M. , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y. , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y. , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I. , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino ,M., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL Unpublished (2000)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki ,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1. .301
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="E130107M24"
/clone_lib="RIKEN full-length enriched, 0 day neonate eyeball"
/tissue_type="eyeball"
/dev_stage="0 day neonate"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT 55 a 105 c 60 g 81 t
ORIGIN

Query Match 87.6%; Score 18.4; DB 9; Length 301;
Best Local Similarity 95.0%; Pred. No. 4.4e+04;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TCTTTTTTTCTCTTCACAGG 21
||||||||||||||| ||||
Db 177 TCTTTTTTTCTCTTCTCAGG 196

RESULT 5
BB125849
LOCUS BB125849 327 bp mRNA linear EST 28-JUN-2000
DEFINITION BB125849 RIKEN full-length enriched, 16 days neonate cerebellum Mus musculus cDNA clone 9630009E03 3', mRNA sequence.
ACCESSION BB125849
VERSION BB125849.1 GI:8780181
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 327)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M. , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y. , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y. , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I. , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino ,M., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL Unpublished (2000)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermolabile enzymes by

trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

```
FEATURES             Location/Qualifiers
     source          1. .327
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="9630009E03"
                     /clone_lib="RIKEN full-length enriched, 16 days neonate
                     cerebellum"
                     /tissue_type="cerebellum"
                     /dev_stage="16 days neonate"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 20.0 and subtraction to Rot = 370.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC
                     3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I."
BASE COUNT       92 a     52 c     49 g    134 t
ORIGIN
```

```
  Query Match             87.6%;  Score 18.4;  DB 9;  Length 327;
  Best Local Similarity   95.0%;  Pred. No. 4.3e+04;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 TCTTTTTTTCTCTTCACAGG 21
          ||||||||||||||| ||||
Db     19 TCTTTTTTTCTCTTCCCAGG 38
```

```
RESULT   6
AA063675/c
LOCUS       AA063675                335 bp    mRNA    linear   EST 24-SEP-1996
DEFINITION  T3357 MVAT4 bloodstream form of serodeme WRATat1.1 Trypanosoma
            brucei rhodesiense cDNA 5', mRNA sequence.
ACCESSION   AA063675
VERSION     AA063675.1  GI:1557735
KEYWORDS    EST.
SOURCE      Trypanosoma brucei rhodesiense.
  ORGANISM  Trypanosoma brucei rhodesiense
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 335)
  AUTHORS   Djikeng,A., Donelson,J.E. and Majiwa,P.A.O.
  TITLE     Generation of expressed sequence tags as physical landmarks in the
            genome of Trypanosoma brucei
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Majiwa PAO
            Molecular Biology Unit
            International Livestock Research Institute
            P.O. Box 30709, Nairobi, Kenya
            Tel: 254-2 630743
            Fax: 254-2 631499
            Email: p.majiwa@cgnet.com
            Seq primer: T3 primer.
FEATURES             Location/Qualifiers
     source          1. .335
                     /organism="Trypanosoma brucei rhodesiense"
                     /db_xref="taxon:31286"
                     /clone_lib="MVAT4 bloodstream form of serodeme WRATat1.1"
                     /note="Vector: Lambda ZAP II (Stratagene); Site_1: EcorI;
                     Site_2: XhoI; The mRNA was purified from a cloned
                     population of bloodstream trypanosomes reexpressing the
                     MVAT4 metacyclic variant surface glycoprotein (VSG). A
                     unidirectional oligo dT-primed EcoRI/XhoI cDNA library was
                     constructed in lambda ZAP II (Stratagene)."
BASE COUNT      115 a     92 c     56 g     72 t
ORIGIN
```

```
  Query Match             87.6%;  Score 18.4;  DB 9;  Length 335;
  Best Local Similarity   95.0%;  Pred. No. 4.2e+04;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TTCTTTTTTTCTCTTCACAG 20
          || |||||||||||||||||
Db    223 TTTTTTTTTCTCTTCACAG 204
```

```
RESULT   7
AQ908298
LOCUS       AQ908298                341 bp    DNA     linear   GSS 09-JAN-2001
DEFINITION  GSSTc05056 Trypanosoma cruzi random genomic library Trypanosoma
            cruzi genomic clone G26J5, DNA sequence.
ACCESSION   AQ908298
VERSION     AQ908298.3  GI:10133097
KEYWORDS    GSS.
SOURCE      Trypanosoma cruzi.
  ORGANISM  Trypanosoma cruzi
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma; Schizotrypanum.
REFERENCE   1  (bases 1 to 341)
  AUTHORS   Aguero,F., Verdun,R., Frasch,A.C.C. and Sanchez,D.O.
  TITLE     A random sequencing approach for the analysis of the trypanosoma
            cruzi genome: general structure, large gene and repetitive DNA
            families, and gene discovery
  JOURNAL   Genome Res. 10 (12), 1996-2005 (2000)
  MEDLINE   20568489
COMMENT     On Sep 14, 2000 this sequence version replaced gi:9370869.
            Contact: Sanchez D.O.
            Instituto de Investigaciones Biotecnologicas (Univ. Nac. de Gral
            San Martin)
            Av. Gral Paz entre Albarellos y Constituyentes, INTI edificio 24
            CP(1650) San Martin, Prov. de BS AS. Argentina
            Tel: 54-11-4580-7255 ext 309
            Fax: 54-11-4752-9639
            Email: dsanchez@iib.unsam.edu.ar
            Sequences were basecalled with phred and vector was masked with
            crossmatch (see http://genome.washington.edu). Sequences were then
            trimmed from both ends to remove low quality bases and masked
            vector.
            Seq primer: T7
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .341
                     /organism="Trypanosoma cruzi"
                     /strain="CL-Brener"
                     /db_xref="taxon:5693"
                     /clone="G26J5"
                     /clone_lib="Trypanosoma cruzi random genomic library"
                     /cell_type="epimastigote"
                     /note="Vector: pBS(-) (Stratagene); T. cruzi DNA was
                     randomly sheared using a nebulizer and the 1 to 2 Kb range
                     was gel purified and cloned into the dephosphoryated
```

```
                HincII site of the vector"
BASE COUNT      108 a     76 c     52 g    105 t
ORIGIN


  Query Match            87.6%;  Score 18.4;  DB 12;  Length 341;
  Best Local Similarity   95.0%;  Pred. No. 4.2e+04;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TTCTTTTTTTCTCTTCACAG 20
          || ||||||||||||||||
Db    148 TTTTTTTTTTCTCTTCACAG 167


RESULT   8
BH057848
LOCUS       BH057848                 349 bp    DNA     linear   GSS 18-JUL-2001
DEFINITION  RPCI-24-354M18.TJ RPCI-24 Mus musculus genomic clone RPCI-24-354M18
            , DNA sequence.
ACCESSION   BH057848
VERSION     BH057848.1  GI:14866229
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 349)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-354M18.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 354  row: M  column: 18
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .349
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-354M18"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT       89 a     89 c     68 g    103 t
ORIGIN


  Query Match            87.6%;  Score 18.4;  DB 12;  Length 349;
  Best Local Similarity   95.0%;  Pred. No. 4.2e+04;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TTCTTTTTTTCTCTTCACAG 20
          || ||||||||||||||||
Db    250 TTTTTTTTTTCTCTTCACAG 269


RESULT   9
AW493746
LOCUS       AW493746                 407 bp    mRNA    linear   EST 24-FEB-2000
DEFINITION  UI-M-BH3-auo-g-01-0-UI.s1 NIH_BMAP_M_S4 Mus musculus cDNA clone
            UI-M-BH3-auo-g-01-0-UI 3', mRNA sequence.
ACCESSION   AW493746
VERSION     AW493746.1  GI:7064027
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 407)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            normalized hypothalamus library cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: Researchers may obtain BMAP cDNA
            clones from RESEARCH GENETICS. It should be noted that Bento Soares
            is generating a small number of additional specialized
            non-redundant arrays of BMAP cDNAs whose availability will be
            considered under appropriate and limited collaborative arrangements
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .407
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-BH3-auo-g-01-0-UI"
                     /clone_lib="NIH_BMAP_M_S4"
                     /dev_stage="27-32 days"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_M_S4 library is a subtracted library of a series,
                     ultimately  derived from a mixture of individually tagged
                     normalized libraries from  ten regions of the mouse brain
                     (cerebellum, brain stems, olfactory  bulbs, hypothalamus,
                     cortex, amygdala, basal ganglia, pineal gland,  striatum,
                     hipoccampus) after a series of subtractions to reduce the
                     representation of cDNAs from which ESTs had already been
                     generated. The following serially subtracted libraries
                     were generated in this process: NIH_BMAP_M_S4,
                     NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1,
                     NIH_BMAP_M_S2, NIH_BMAP_M_S1.  The subtracted library
                     (NIH_BMAP_M_S4) was constructed as follows:  PCRamplified
                     cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and
                     NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived
                     was used as  a driver in a hybridization with a pool of
                     the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1
                     libraries in the form of  single-stranded circles.  The
                     remaining single-stranded circles  (subtracted library)
                     was purified by hydroxyapatite column chromatography,
                     converted to double-stranded circles and electroporated
                     into DH10B  bacteria (LifeTechnologies) to generate the
                     NIH_BMAP_M_S4 library.  This procedure has been previously
```

```
                described (Bonaldo, Lennon and Soares, Genome Research
                6:791-806, 1996)
                TAG_LIB=NIH_BMAP_M_S4
                TAG_TISSUE=hypothalamus
                TAG_SEQ=CGGCA"
BASE COUNT       94 a    102 c     70 g    141 t
ORIGIN


  Query Match             87.6%;  Score 18.4;  DB 9;  Length 407;
  Best Local Similarity   95.0%;  Pred. No. 3.9e+04;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TTCTTTTTTTCTCTTCACAG 20
          || ||||||||||||||||
Db      6 TTTTTTTTTTCTCTTCACAG 25


RESULT 10
BG140084/c
LOCUS       BG140084                511 bp    mRNA    linear   EST 31-JAN-2001
DEFINITION  EST480526 wild tomato pollen Lycopersicon pennellii cDNA clone
            cLPP16I19 5' sequence, mRNA sequence.
ACCESSION   BG140084
VERSION     BG140084.1  GI:12640272
KEYWORDS    EST.
SOURCE      Lycopersicon pennellii.
  ORGANISM  Lycopersicon pennellii
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon.
REFERENCE   1  (bases 1 to 511)
  AUTHORS   van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Utterback,T.,
            Hansen,C., Ronning,C. and Tanksley,S.
  TITLE     Generation of ESTs from wild tomato (L. pennellii) pollen
  JOURNAL   Unpublished (2001)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html.
FEATURES             Location/Qualifiers
     source          1. .511
                     /organism="Lycopersicon pennellii"
                     /cultivar="TA56"
                     /db_xref="taxon:28526"
                     /clone="cLPP16I19"
                     /clone_lib="wild tomato pollen"
                     /tissue_type="pollen"
                     /dev_stage="pollen collected from open flowers"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript SK(-); Site_1: EcoR1; Site_2:
                     Xho1; Pollen was collected from open flowers from
                     L.pennellii TA56, and stored at -80 C until library
                     construction."
BASE COUNT      156 a     77 c    125 g    153 t
ORIGIN


  Query Match             87.6%;  Score 18.4;  DB 10;  Length 511;
  Best Local Similarity   95.0%;  Pred. No. 3.6e+04;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TTCTTTTTTTCTCTTCACAG 20
          |||||| ||||||||||||||
Db    432 TTCTTTCTTTCTCTTCACAG 413


RESULT 11
AA677704/c
LOCUS       AA677704                522 bp    mRNA    linear   EST 19-DEC-1997
DEFINITION  zj72h09.s1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA
            clone IMAGE:460481 3', mRNA sequence.
ACCESSION   AA677704
VERSION     AA677704.1  GI:2658226
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 522)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
            ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 468.
FEATURES             Location/Qualifiers
     source          1. .522
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:460481"
                     /clone_lib="Soares_fetal_liver_spleen_1NFLS_S1"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     This is a subtracted version of the original Soares fetal
                     liver spleen 1NFLS library.  1st strand cDNA was primed
                     with a Pac I - oligo(dT) primer [5'
                     AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      186 a    100 c    109 g    127 t
ORIGIN


  Query Match             87.6%;  Score 18.4;  DB 9;  Length 522;
  Best Local Similarity   95.0%;  Pred. No. 3.5e+04;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TTCTTTTTTTCTCTTCACAG 20
          |||||| ||||||||||||||
Db    472 TTCTTTATTTCTCTTCACAG 453


RESULT 12
BM217071
LOCUS       BM217071                561 bp    mRNA    linear   EST 31-JAN-2002
DEFINITION  C0890E05-3 NIA Mouse Blastocyst cDNA Library (Long) Mus musculus
            cDNA clone C0890E05 3', mRNA sequence.
ACCESSION   BM217071
VERSION     BM217071.1  GI:17776109
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 561)
  AUTHORS   Piao,Y., Kargul,G.J., Dudekula,D.B., Qian,Y., Tanaka,T., Luo,A. and
            Ko,M.S.H.
```

```
  TITLE     Systematic Analyses of NIA Mouse Blastocyst cDNA Library (Long)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Dawood B. Dudekula
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdna@lgsun.grc.nia.nih.gov
            Plate: C0890  row: E  column: 05
            Seq primer: -21M13 Forward
            High quality sequence stop: 561
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .561
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="niaEST:C0890E05-3"
                     /db_xref="taxon:10090"
                     /clone="C0890E05"
                     /clone_lib="NIA Mouse Blastocyst cDNA Library (Long)"
                     /tissue_type="Blastocyst"
                     /dev_stage="3.5-dpc"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2:
                     NotI; Mouse cDNA project by the Laboratory of Genetics,
                     National Institute on Aging (NIA), Intramural Research
                     Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is
                     a long-transcript enriched cDNA library (Ref. Genome Res.
                     11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were
                     extracted from a pool of 20 Blastocysts. Double-stranded
                     cDNAs were synthesized with an Oligo(dT) primer
                     [Invitrogen:
                     5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTTT-3'] from
                     0.2 ug of total RNA, treated with T4 DNA polymerase, and
                     purified by ethanol-precipitation. The cDNAs were ligated
                     to Lone-linker LL-Sal4, purified by phenol/chloroform, and
                     separated from free linkers by Centricon 100. Then, the
                     cDNAs were amplified by long-range high fidelity PCR using
                     Ex Taq polymerase (Takara) with a primer Sal4-S. The
                     products were purified by phenol/chloroform and Centricon
                     100. The cDNAs were digested with SalI and NotI enzymes
                     and cloned into SalI/NotI site of pSPORT1 plasmid vector.
                     The DH10B E. coli host was transformed with the ligation
                     mixture by the standard chemical method. The average
                     insert size is about 2.2 kb. The library was constructed
                     by Yulan Piao (NIA)."
BASE COUNT      145 a    127 c    127 g    162 t
ORIGIN


  Query Match             87.6%;  Score 18.4;  DB 10;  Length 561;
  Best Local Similarity   95.0%;  Pred. No. 3.4e+04;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TTCTTTTTTTCTCTTCACAG 20
          |||||||||| |||||||||
Db    114 TTCTTTTTTTATCTTCACAG 133


RESULT  13
BM030926/c
LOCUS       BM030926                 577 bp    mRNA    linear   EST 05-NOV-2001
DEFINITION  495476 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BM030926
VERSION     BM030926.1  GI:16744496
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 577)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 122  row: D  column: 15
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .577
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 2BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from testis, thymus,
                     semitendonosus muscle, longissimus muscle, pancreas,
                     adrenal, and endometrium."
BASE COUNT      145 a    121 c    144 g    167 t
ORIGIN


  Query Match             87.6%;  Score 18.4;  DB 10;  Length 577;
  Best Local Similarity   95.0%;  Pred. No. 3.4e+04;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 TCTTTTTTTCTCTTCACAGG 21
          |||||||||| ||||||||||
Db    141 TCTTTTTTTTTCTTCACAGG 122


RESULT  14
AI671885/c
LOCUS       AI671885                 593 bp    mRNA    linear   EST 17-DEC-1999
DEFINITION  wb41b12.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2308223 3',
            mRNA sequence.
ACCESSION   AI671885
VERSION     AI671885.1  GI:4851616
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 593)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
```

```
           Insert Length: 744   Std Error: 0.00
           Seq primer: -40UP from Gibco
           High quality sequence stop: 450.
FEATURES             Location/Qualifiers
     source          1. .593
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2308223"
                     /clone_lib="NCI_CGAP_GC6"
                     /tissue_type="pooled germ cell tumors"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA
                     from the normalized library NCI_CGAP_GC4 was prepared, and
                     ss circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from a pool
                     of 5,000 clones made from the same library (cloneIDs
                     1257096-1258631, 1469064-1470983, and 1475592-1476743).
                     Subtraction by Bento Soares and M. Fatima Bonaldo.  "
BASE COUNT      213 a    109 c    122 g    148 t      1 others
ORIGIN


  Query Match             87.6%;  Score 18.4;  DB 9;  Length 593;
  Best Local Similarity   95.0%;  Pred. No. 3.2e+04;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TTCTTTTTTTCTCTTCACAG 20
          |||||| ||||||||||||||
Db    477 TTCTTTATTTCTCTTCACAG 458


RESULT  15
AG055132/c
LOCUS       AG055132                679 bp    DNA     linear   GSS 02-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-041A09.R, genomic survey sequence.
ACCESSION   AG055132
VERSION     AG055132.1  GI:16592575
KEYWORDS    GSS; GSS (genome survey sequence).
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-041A09.R.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (sites)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 679)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: M13Rev
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .679
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-041A09.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      147 a    170 c    163 g    198 t      1 others
ORIGIN


  Query Match             87.6%;  Score 18.4;  DB 12;  Length 679;
  Best Local Similarity   95.0%;  Pred. No. 3.2e+04;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 TCTTTTTTTCTCTTCACAGG 21
          |||| ||||||||||||||
Db    245 TCTTATTTTCTCTTCACAGG 226


Search completed: July 21, 2002, 09:11:09
Job time: 10385 sec
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2002, 06:20:09 ; Search time 2038.31 Seconds
(without alignments)
92.400 Million cell updates/sec

Title: US-09-754-014-10_COPY_1_9
Perfect score: 9
Sequence: 1 CAGGTAAGT 9

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	9	100.0	10	6	AX152396	AX152396 Sequence
2	9	100.0	13	6	AX018745	AX018745 Sequence
c 3	9	100.0	13	6	AX018746	AX018746 Sequence
c 4	9	100.0	14	6	A63967	A63967 Sequence 11
c 5	9	100.0	14	6	AR091477	AR091477 Sequence
6	9	100.0	14	6	AX018747	AX018747 Sequence
c 7	9	100.0	14	6	AX018748	AX018748 Sequence
8	9	100.0	15	6	AX018749	AX018749 Sequence
c 9	9	100.0	15	6	AX018750	AX018750 Sequence
10	9	100.0	20	6	AR075705	AR075705 Sequence
11	9	100.0	20	6	AR139514	AR139514 Sequence
12	9	100.0	20	6	AX069132	AX069132 Sequence
13	9	100.0	20	6	AX259809	AX259809 Sequence
14	9	100.0	20	6	E15758	E15758 PCR primer
15	9	100.0	20	6	E21777	E21777 Novel slit-
16	9	100.0	20	6	E22925	E22925 Slit-like p
17	9	100.0	20	6	E26717	E26717 Vertebrate
18	9	100.0	20	6	I85583	I85583 Sequence 4
c 19	9	100.0	21	6	AR072471	AR072471 Sequence
20	9	100.0	22	4	DOGP41302	L24301 Dog (Clone:
c 21	9	100.0	24	6	AX321460	AX321460 Sequence
22	9	100.0	24	6	E36207	E36207 Process for
23	9	100.0	24	6	E37235	E37235 Method for
24	9	100.0	24	6	E47020	E47020 Simulataneo
c 25	9	100.0	25	6	AX115172	AX115172 Sequence
26	9	100.0	26	6	AX081476	AX081476 Sequence
27	9	100.0	28	6	AR090890	AR090890 Sequence
28	9	100.0	28	6	AX277381	AX277381 Sequence
29	9	100.0	30	6	I36150	I36150 Sequence 34
30	9	100.0	31	6	A84890	A84890 Sequence 39
31	9	100.0	31	6	AX211662	AX211662 Sequence
c 32	9	100.0	31	6	BD002353	BD002353 Gene comp
33	9	100.0	32	6	AX057068	AX057068 Sequence
c 34	9	100.0	32	6	AX183886	AX183886 Sequence
35	9	100.0	35	6	AX338460	AX338460 Sequence
c 36	9	100.0	36	6	AX139056	AX139056 Sequence
c 37	9	100.0	36	6	AX139058	AX139058 Sequence
c 38	9	100.0	38	6	AX018808	AX018808 Sequence
39	9	100.0	39	6	AR052806	AR052806 Sequence
40	9	100.0	39	6	AR127978	AR127978 Sequence
41	9	100.0	40	6	A83613	A83613 Sequence 42
42	9	100.0	41	6	AR176210	AR176210 Sequence
c 43	9	100.0	42	6	AR030312	AR030312 Sequence
c 44	9	100.0	43	6	AX079954	AX079954 Sequence
45	9	100.0	43	6	AX079955	AX079955 Sequence

ALIGNMENTS

RESULT 1
AX152396/c
LOCUS AX152396 10 bp DNA linear PAT 22-JUN-2001
DEFINITION Sequence 311 from Patent WO0138577.
ACCESSION AX152396
VERSION AX152396.1 GI:14534047
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 10)
AUTHORS Velculescu,V.E., Vogelstein,B. and Kinzler,K.W.
TITLE Human transcriptomes
JOURNAL Patent: WO 0138577-A 311 31-MAY-2001;
The Johns Hopkins University (US)
FEATURES Location/Qualifiers
source 1. .10
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 2 a 4 c 1 g 3 t
ORIGIN

```
  Query Match           100.0%;  Score 9;  DB 6;  Length 10;
  Best Local Similarity  100.0%;  Pred. No. 6.5e+04;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CAGGTAAGT 9
          |||||||||
Db      9 CAGGTAAGT 1


RESULT   2
AX018745
LOCUS       AX018745                 13 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 3 from Patent WO9943848.
ACCESSION   AX018745
VERSION     AX018745.1  GI:10042868
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 13)
  AUTHORS   Ong,C.J. and Jirik,F.R.
  TITLE     Protein interaction and transcription factor trap
  JOURNAL   Patent: WO 9943848-A 3 02-SEP-1999;
            ONG CHRISTOPHER J (CA); UNIV BRITISH COLUMBIA (CA); JIRIK FRANK R
            (CA)
FEATURES             Location/Qualifiers
     source          1. .13
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="Oligomer containing a splice donor sequence"
BASE COUNT        4 a      2 c      4 g      3 t
ORIGIN


  Query Match           100.0%;  Score 9;  DB 6;  Length 13;
  Best Local Similarity  100.0%;  Pred. No. 6.3e+04;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CAGGTAAGT 9
          |||||||||
Db      5 CAGGTAAGT 13


RESULT   3
AX018746/c
LOCUS       AX018746                 13 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 4 from Patent WO9943848.
ACCESSION   AX018746
VERSION     AX018746.1  GI:10042869
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 13)
  AUTHORS   Ong,C.J. and Jirik,F.R.
  TITLE     Protein interaction and transcription factor trap
  JOURNAL   Patent: WO 9943848-A 4 02-SEP-1999;
            ONG CHRISTOPHER J (CA); UNIV BRITISH COLUMBIA (CA); JIRIK FRANK R
            (CA)
FEATURES             Location/Qualifiers
     source          1. .13
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="Oligomer for adjusting a reading frame for
                     ligation"
BASE COUNT        3 a      4 c      1 g      5 t
ORIGIN


  Query Match           100.0%;  Score 9;  DB 6;  Length 13;
  Best Local Similarity  100.0%;  Pred. No. 6.3e+04;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CAGGTAAGT 9
          |||||||||
Db     13 CAGGTAAGT 5


RESULT   4
A63967/c
LOCUS       A63967                   14 bp    DNA     linear   PAT 29-MAR-1999
DEFINITION  Sequence 11 from Patent EP0784094.
ACCESSION   A63967
VERSION     A63967.1  GI:3717488
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 14)
  AUTHORS   Bozzoni,I.
  TITLE     Ribozyme -snRNA chimeric molecules having a catalytic activity for
            nuclear RNAs
  JOURNAL   Patent: EP 0784094-A 11 16-JUL-1997;
            UNIV ROMA (IT)
FEATURES             Location/Qualifiers
     source          1. .14
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        3 a      4 c      3 g      4 t
ORIGIN


  Query Match           100.0%;  Score 9;  DB 6;  Length 14;
  Best Local Similarity  100.0%;  Pred. No. 6.3e+04;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CAGGTAAGT 9
          |||||||||
Db     12 CAGGTAAGT 4


RESULT   5
AR091477/c
LOCUS       AR091477                 14 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 11 from patent US 5994124.
ACCESSION   AR091477
VERSION     AR091477.1  GI:10018232
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 14)
  AUTHORS   Bozzoni,I.
  TITLE     Ribozyme-snRNA chimeric molecules having a catalytic activity for
            nuclear RNAs
  JOURNAL   Patent: US 5994124-A 11 30-NOV-1999;
FEATURES             Location/Qualifiers
     source          1. .14
                     /organism="unknown"
BASE COUNT        3 a      4 c      3 g      4 t
ORIGIN


  Query Match           100.0%;  Score 9;  DB 6;  Length 14;
  Best Local Similarity  100.0%;  Pred. No. 6.3e+04;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CAGGTAAGT 9
          |||||||||
Db     12 CAGGTAAGT 4


RESULT   6
AX018747
LOCUS       AX018747                 14 bp    DNA     linear   PAT 07-SEP-2000
```

```
DEFINITION  Sequence 5 from Patent WO9943848.
ACCESSION   AX018747
VERSION     AX018747.1  GI:10042870
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 14)
  AUTHORS   Ong,C.J. and Jirik,F.R.
  TITLE     Protein interaction and transcription factor trap
  JOURNAL   Patent: WO 9943848-A 5 02-SEP-1999;
            ONG CHRISTOPHER J (CA); UNIV BRITISH COLUMBIA (CA); JIRIK FRANK R
            (CA)
FEATURES             Location/Qualifiers
     source          1. .14
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="Oligomer containing a splice donor sequence"
BASE COUNT        4 a      3 c      4 g      3 t
ORIGIN


  Query Match             100.0%;  Score 9;  DB 6;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 6.3e+04;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CAGGTAAGT 9
          |||||||||
Db      6 CAGGTAAGT 14


RESULT   7
AX018748/c
LOCUS       AX018748                  14 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 6 from Patent WO9943848.
ACCESSION   AX018748
VERSION     AX018748.1  GI:10042871
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 14)
  AUTHORS   Ong,C.J. and Jirik,F.R.
  TITLE     Protein interaction and transcription factor trap
  JOURNAL   Patent: WO 9943848-A 6 02-SEP-1999;
            ONG CHRISTOPHER J (CA); UNIV BRITISH COLUMBIA (CA); JIRIK FRANK R
            (CA)
FEATURES             Location/Qualifiers
     source          1. .14
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="Oligomer for adjusting a reading frame for
                     ligation"
BASE COUNT        3 a      4 c      2 g      5 t
ORIGIN


  Query Match             100.0%;  Score 9;  DB 6;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 6.3e+04;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CAGGTAAGT 9
          |||||||||
Db     13 CAGGTAAGT 5


RESULT   8
AX018749
LOCUS       AX018749                  15 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 7 from Patent WO9943848.
ACCESSION   AX018749
VERSION     AX018749.1  GI:10042872
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 15)
  AUTHORS   Ong,C.J. and Jirik,F.R.
  TITLE     Protein interaction and transcription factor trap
  JOURNAL   Patent: WO 9943848-A 7 02-SEP-1999;
            ONG CHRISTOPHER J (CA); UNIV BRITISH COLUMBIA (CA); JIRIK FRANK R
            (CA)
FEATURES             Location/Qualifiers
     source          1. .15
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="Oligomer containing a splice donor sequence"
BASE COUNT        4 a      4 c      4 g      3 t
ORIGIN


  Query Match             100.0%;  Score 9;  DB 6;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 6.2e+04;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CAGGTAAGT 9
          |||||||||
Db      7 CAGGTAAGT 15


RESULT   9
AX018750/c
LOCUS       AX018750                  15 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 8 from Patent WO9943848.
ACCESSION   AX018750
VERSION     AX018750.1  GI:10042873
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 15)
  AUTHORS   Ong,C.J. and Jirik,F.R.
  TITLE     Protein interaction and transcription factor trap
  JOURNAL   Patent: WO 9943848-A 8 02-SEP-1999;
            ONG CHRISTOPHER J (CA); UNIV BRITISH COLUMBIA (CA); JIRIK FRANK R
            (CA)
FEATURES             Location/Qualifiers
     source          1. .15
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="Oligomer for adjusting a reading frame for
                     ligation"
BASE COUNT        3 a      4 c      3 g      5 t
ORIGIN


  Query Match             100.0%;  Score 9;  DB 6;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 6.2e+04;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CAGGTAAGT 9
          |||||||||
Db     13 CAGGTAAGT 5


RESULT  10
AR075705
LOCUS       AR075705                  20 bp    DNA     linear   PAT 30-AUG-2000
DEFINITION  Sequence 4 from patent US 5958692.
ACCESSION   AR075705
VERSION     AR075705.1  GI:10002451
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 20)
```

```
  AUTHORS   Cotton,R.G.H., Youil,R. and Kemper,B.W.
  TITLE     Detection of mutation by resolvase cleavage
  JOURNAL   Patent: US 5958692-A 4 28-SEP-1999;
FEATURES             Location/Qualifiers
     source          1. .20
                     /organism="unknown"
BASE COUNT        5 a      4 c      7 g      4 t
ORIGIN


  Query Match            100.0%;  Score 9;  DB 6;  Length 20;
  Best Local Similarity  100.0%;  Pred. No. 6e+04;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CAGGTAAGT 9
          |||||||||
Db      2 CAGGTAAGT 10


RESULT  11
AR139514
LOCUS       AR139514                  20 bp    DNA     linear   PAT 16-JUN-2001
DEFINITION  Sequence 31 from patent US 6207383.
ACCESSION   AR139514
VERSION     AR139514.1  GI:14482010
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 20)
  AUTHORS   Keating,M.T. and Splawski,I.
  TITLE     Mutations in and genomic structure of HERG--a long QT syndrome gene
  JOURNAL   Patent: US 6207383-A 31 27-MAR-2001;
FEATURES             Location/Qualifiers
     source          1. .20
                     /organism="unknown"
BASE COUNT        4 a      6 c      6 g      4 t
ORIGIN


  Query Match            100.0%;  Score 9;  DB 6;  Length 20;
  Best Local Similarity  100.0%;  Pred. No. 6e+04;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CAGGTAAGT 9
          |||||||||
Db      8 CAGGTAAGT 16


RESULT  12
AX069132
LOCUS       AX069132                  20 bp    DNA     linear   PAT 25-JAN-2001
DEFINITION  Sequence 50 from Patent WO0102604.
ACCESSION   AX069132
VERSION     AX069132.1  GI:12579014
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 20)
  AUTHORS   Tournier-Lasserve,E., Laberge-Le,S. and Labauge,P.
  TITLE     Use of the krit1 gene in angiogenesis
  JOURNAL   Patent: WO 0102604-A 50 11-JAN-2001;
            INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE (INSERM)
            (FR)
FEATURES             Location/Qualifiers
     source          1. .20
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT       10 a      3 c      3 g      4 t
ORIGIN


  Query Match            100.0%;  Score 9;  DB 6;  Length 20;
  Best Local Similarity  100.0%;  Pred. No. 6e+04;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CAGGTAAGT 9
          |||||||||
Db      8 CAGGTAAGT 16


RESULT  13
AX259809
LOCUS       AX259809                  20 bp    DNA     linear   PAT 26-OCT-2001
DEFINITION  Sequence 36 from Patent WO0172822.
ACCESSION   AX259809
VERSION     AX259809.1  GI:16508883
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Hugot,J.P., Thomas,G., Zouali,M., Lesage,S. and Chamaillard,M.
  TITLE     Genes involved in intestinal inflammatory diseases and use thereof
  JOURNAL   Patent: WO 0172822-A 36 04-OCT-2001;
            Fondation Jean Dausset-Ceph (FR)
FEATURES             Location/Qualifiers
     source          1. .20
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT        6 a      3 c      6 g      5 t
ORIGIN


  Query Match            100.0%;  Score 9;  DB 6;  Length 20;
  Best Local Similarity  100.0%;  Pred. No. 6e+04;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CAGGTAAGT 9
          |||||||||
Db     10 CAGGTAAGT 18


RESULT  14
E15758
LOCUS       E15758                    20 bp    DNA     linear   PAT 28-JUL-1999
DEFINITION  PCR primer for human Slit cDNA.
ACCESSION   E15758
VERSION     E15758.1  GI:5710441
KEYWORDS    JP 1998087699-A/5.
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 20)
  AUTHORS   Ito,A. and Sakano,S.
  TITLE     SLIT-LIKE POLYPEPTIDE
  JOURNAL   Patent: JP 1998087699-A 5 07-APR-1998;
            ASAHI CHEM IND CO LTD
COMMENT     OS   None
            OC   Artificial sequences.
            PN   JP 1998087699-A/5
            PD   07-APR-1998
            PF   15-JUL-1997 JP 1997205351
            PR   16-JUL-1996 JP 96P   186219
            PI   ITO AKIRA, SAKANO SEIJI
            PC   C07K14/47,A61K38/00,C07K16/18,C12N5/10,C12N15/09,C12N15/02, PC
              C12P21/02,
            PC   C12P21/08,(C12P21/02,C12R1:91);
            CC   strandedness: Single;
            CC   topology: Linear;
            FH   Key             Location/Qualifiers
            FH
            FT   source          1. .20
```

```
           FT                   /organism='Artificial sequences'.
FEATURES             Location/Qualifiers
     source          1. .20
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        5 a      4 c      6 g      5 t
ORIGIN


  Query Match             100.0%;  Score 9;  DB 6;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 6e+04;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CAGGTAAGT 9
          |||||||||
Db      8 CAGGTAAGT 16


RESULT  15
E21777
LOCUS       E21777                    20 bp    DNA     linear   PAT 07-FEB-2001
DEFINITION  Novel slit-like polypeptide.
ACCESSION   E21777
VERSION     E21777.1  GI:13023697
KEYWORDS    JP 1999018777-A/8.
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 20)
  AUTHORS   Akira,I.S.S.S.
  TITLE     Novel slit-like polypeptide
  JOURNAL   Patent: JP 1999018777-A 8 26-JAN-1999;
            ASAHI CHEM IND CO LTD
COMMENT     OS   Unidentified
            PN   JP 1999018777-A/8
            PD   26-JAN-1999
            PF   09-JUL-1997 JP 1997183683
            PR
            PI   AKIRA ITO,SEIJI SAKANO
            PC   C12N15/09,A61K38/00,A61K38/00,A61K38/00,C07H21/04,C07K14/47,
            PC   C07K16/18,
            PC   C12N5/10,C12P21/02//(C12N15/09,C12R1:91),(C12N5/10,C12R1:91),
            PC   (C12P21/02,C12R1:91),C12N15/00,A61K37/02,A61K37/02,A61K37/02,
            PC   C12N5/00,
            PC   (C12N15/00,C12R1:91),(C12N5/00,C12R1:91)
            CC   Strandedness: Single;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers
            FT   source          1. .20
            FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .20
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        5 a      4 c      6 g      5 t
ORIGIN


  Query Match             100.0%;  Score 9;  DB 6;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 6e+04;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CAGGTAAGT 9
          |||||||||
Db      8 CAGGTAAGT 16


Search completed: July 21, 2002, 09:45:13
Job time: 12304 sec
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2002, 08:08:59 ; Search time 467.25 Seconds
(without alignments)
33.071 Million cell updates/sec

Title: US-09-754-014-10_COPY_1_9
Perfect score: 9
Sequence: 1 CAGGTAAGT 9

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	9	100.0	9	19	AAV43548	Insertion sequence
c	2	9	100.0	10	21	AAZ81555	Metastatic breast
c	3	9	100.0	10	21	AAZ82988	Metastatic breast
c	4	9	100.0	10	21	AAZ85454	Metastatic breast
c	5	9	100.0	10	22	AAH43531	SD sequence. Synt
c	6	9	100.0	10	22	AAH63471	Human ubiquitously
	7	9	100.0	11	22	AAH23795	Murine histidine d
	8	9	100.0	13	20	AAZ11271	Splice donor site
	9	9	100.0	14	20	AAZ11272	Splice donor site
	10	9	100.0	15	20	AAZ40412	5' splice site seq
	11	9	100.0	15	20	AAZ11273	Splice donor site
c	12	9	100.0	17	18	AAX71417	Human KDR VEGF rec
c	13	9	100.0	17	18	AAX71418	Human KDR VEGF rec
	14	9	100.0	17	21	AAA36045	Human genomic SNP
c	15	9	100.0	19	21	AAC72504	Single nucleotide
	16	9	100.0	20	16	AAT00608	21-hydroxylase B g
	17	9	100.0	20	19	AAV16969	Oligonucleotide se
	18	9	100.0	20	20	AAX89172	Seq ID No: 26 of J
	19	9	100.0	20	20	AAX19950	Human slit 1 PCR a
	20	9	100.0	20	20	AAX14987	PCR primer used to
	21	9	100.0	20	20	AAV80054	Human PMM2 exon 1/
	22	9	100.0	20	21	AAZ74093	Human biallelic ma
	23	9	100.0	20	21	AAZ74129	Human biallelic ma
c	24	9	100.0	20	21	AAC68759	Human FUT3 antisen
	25	9	100.0	20	21	AAZ93668	Antisense oligonuc
	26	9	100.0	20	21	AAZ93669	Antisense oligonuc
	27	9	100.0	20	21	AAZ93670	Antisense oligonuc
	28	9	100.0	20	21	AAZ93671	Antisense oligonuc
	29	9	100.0	20	21	AAZ93672	Antisense oligonuc
	30	9	100.0	20	21	AAZ93673	Antisense oligonuc
	31	9	100.0	20	21	AAA07629	HERG gene exon 3/i
	32	9	100.0	20	21	AAZ61274	5' splice donor of
	33	9	100.0	20	22	AAI65624	Primer for microsa
	34	9	100.0	20	22	AAS43121	Human ERbeta gene,
	35	9	100.0	20	22	AAS43140	Human ERbeta gene,
	36	9	100.0	20	22	AAS08768	Human PD-ABC form
	37	9	100.0	20	22	AAS08859	Human PD-ABC form
	38	9	100.0	20	22	AAH27713	Human bcl-x splice
	39	9	100.0	20	22	AAH27714	Human bcl-x splice
	40	9	100.0	20	22	AAH27715	Human bcl-x splice
	41	9	100.0	20	22	AAH27716	Human bcl-x splice
	42	9	100.0	20	22	AAH27717	Human bcl-x splice
	43	9	100.0	20	22	AAH27718	Human bcl-x splice
	44	9	100.0	20	22	AAF24987	Nucleotide sequenc
	45	9	100.0	20	24	ABA89794	Human oestrogen re

ALIGNMENTS

RESULT 1
AAV43548
ID AAV43548 standard; DNA; 9 BP.
XX
AC AAV43548;
XX
DT 16-SEP-1998 (first entry)
XX
DE Insertion sequence 1 used for creating a tagged gene.
XX
KW Tagged gene; tagged transcript; hybrid intron; protein tag;
KW protein isolation; recombination; subcellular structure analysis;
KW transcriptional regulation; viral infection; ss.
XX
OS Synthetic.
OS Unidentified.
XX
PN WO9820031-A1.
XX
PD 14-MAY-1998.
XX
PF 07-NOV-1997; 97WO-US20150.
XX
PR 08-NOV-1996; 96US-0705404.
XX
PA (JARV/) JARVIK J W.
XX
PI Jarvik JW;
XX
DR WPI; 1998-286861/25.
XX
PT Tagging genes, transcripts and proteins - using tag-creating DNA

PT inserted into intron of gene to create 2 hybrid introns separated by
PT new exon encoding protein tag
XX
PS Claim 1; Page 33; 66pp; English.
XX
CC This sequence is used in the method of invention for tagging genes,
CC transcripts and proteins in cells in a single recombinational event. The
CC method comprises producing a tagged gene by inserting a DNA sequence
CC into an intron of a gene by selecting a DNA sequence having a 5' portion
CC free of any nucleotide sequence selected from AAV43548 to AAV43551, a
CC nucleotide sequence selected from AAV43552 to AAV43560 and nucleotide
CC sequences identical to a known splice branch site in a known gene,
CC sequences identical in length to a known spacer region between splice
CC branch and acceptor sites in a known gene, sequences identical to a known
CC splice acceptor site in a known gene, sequence identical to a known
CC splice donor site in a known gene, an open reading frame (ORF) 3N-1
CC nucleotides in length, the ORF encoding a known peptide tag recognisable
CC by a known reaction characteristic of the known peptide tag and sequences
CC selected form CAGG and TAGG. The DNA sequence is inserted into the intron
CC within the gene to create a tagged gene, and the tagged gene is incubated
CC within a cell so as to maintain intact or to introduce the tagged gene
CC within the genome of the cell. The method is used for isolating proteins,
CC RNA and genes, for analysis of subcellular structures, cellular responses
CC and transcriptional regulation, for the study of viral infection and for
CC diagnosis of disease.
XX
SQ Sequence 9 BP; 3 A; 1 C; 3 G; 2 T; 0 other;

Query Match 100.0%; Score 9; DB 19; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.9e+08;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAGGTAAGT 9
|||||||||
Db 1 caggtaagt 9

RESULT 2
AAZ81555/c
ID AAZ81555 standard; DNA; 10 BP.
XX
AC AAZ81555;
XX
DT 07-APR-2000 (first entry)
XX
DE Metastatic breast tumour cell upregulated transcript tag #789.
XX
KW Human; metastatic breast tumour tissue; breast cancer; tag; primer;
KW non-metastatic breast tumour tissue; gene therapy; anticancer;
KW antimetastatic; vaccine; diagnosis; ss.
XX
OS Homo sapiens.
XX
PN WO9965928-A2.
XX
PD 23-DEC-1999.
XX
PF 18-JUN-1999; 99WO-US13647.
XX
PR 19-JUN-1998; 98US-0089853.
PR 19-JUN-1998; 98US-0089997.
PR 19-JUN-1998; 98US-0090039.
PR 19-JUN-1998; 98US-0090040.
PR 19-JUN-1998; 98US-0090041.
XX
PA (GENZ) GENZYME CORP.
PA (ROBE/) ROBERTS B L.
PA (SHAN/) SHANKARA S.
XX
PI Roberts BL, Shankara S;
XX
DR WPI; 2000-106079/09.
XX
PT Isolated polynucleotides differentially expressed between metastatic
PT and non-metastatic breast cancer cells, useful for diagnosis,
PT prevention and treatment of cancer -
XX
PS Claim 1; Page 79; 219pp; English.
XX
CC AAZ80767 to AAZ83941 represent tags corresponding to distinct
CC transcripts that are preferentially transcribed in the metastatic breast
CC tumour tissue (i.e. are upregulated in metastatic breast tumour cells).
CC AAZ83942 to AAZ86677 represent tags corresponding to distinct transcripts
CC that are preferentially transcribed in the primary or non-metastatic
CC breast tumour tissue (i.e. are downregulated in metastatic breast tumour
CC cells). These transcripts can be used for diagnosis, prognosis,
CC monitoring and treatment of breast cancer, particularly where metastatic.
CC Diagnosis is by standard immunoassays or hybridisation/amplification
CC reactions. Compounds that modulate expression of the transcripts are
CC potentially useful for treatment of (metastatic) breast cancer, while
CC promoters from the transcripts are used to direct expression, in selected
CC cell types, of e.g. therapeutic genes (also ribozymes or antisense
CC sequences), particularly an antigen-encoding sequence for use in gene or
CC cell-based vaccines. Polypeptides encoded by the transcripts are also
CC useful in vaccines; for diagnosing breast cancer and for raising
CC specific antibodies (Ab). Ab are used to detect the polypeptides or as
CC therapeutic agents. Host cells that produce the polypeptides can be used
CC to expand and isolate populations of educated, antigen-specific immune
CC effecter cells, e.g. cytotoxic T lymphocytes, and these used for
CC adoptive immunotherapy.
XX
SQ Sequence 10 BP; 2 A; 3 C; 2 G; 3 T; 0 other;

Query Match 100.0%; Score 9; DB 21; Length 10;
Best Local Similarity 100.0%; Pred. No. 8.3e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAGGTAAGT 9
|||||||||
Db 9 CAGGTAAGT 1

RESULT 3
AAZ82988/c
ID AAZ82988 standard; DNA; 10 BP.
XX
AC AAZ82988;
XX
DT 07-APR-2000 (first entry)
XX
DE Metastatic breast tumour cell upregulated transcript tag #2222.
XX
KW Human; metastatic breast tumour tissue; breast cancer; tag; primer;
KW non-metastatic breast tumour tissue; gene therapy; anticancer;
KW antimetastatic; vaccine; diagnosis; ss.
XX
OS Homo sapiens.
XX
PN WO9965928-A2.
XX
PD 23-DEC-1999.
XX
PF 18-JUN-1999; 99WO-US13647.
XX
PR 19-JUN-1998; 98US-0089853.
PR 19-JUN-1998; 98US-0089997.
PR 19-JUN-1998; 98US-0090039.
PR 19-JUN-1998; 98US-0090040.
PR 19-JUN-1998; 98US-0090041.
XX
PA (GENZ) GENZYME CORP.
PA (ROBE/) ROBERTS B L.
PA (SHAN/) SHANKARA S.
XX

PI Roberts BL, Shankara S;
XX
DR WPI; 2000-106079/09.
XX
PT Isolated polynucleotides differentially expressed between metastatic
PT and non-metastatic breast cancer cells, useful for diagnosis,
PT prevention and treatment of cancer -
XX
PS Claim 1; Page 119; 219pp; English.
XX
CC AAZ80767 to AAZ83941 represent tags corresponding to distinct
CC transcripts that are preferentially transcribed in the metastatic breast
CC tumour tissue (i.e. are upregulated in metastatic breast tumour cells).
CC AAZ83942 to AAZ86677 represent tags corresponding to distinct transcripts
CC that are preferentially transcribed in the primary or non-metastatic
CC breast tumour tissue (i.e. are downregulated in metastatic breast tumour
CC cells). These transcripts can be used for diagnosis, prognosis,
CC monitoring and treatment of breast cancer, particularly where metastatic.
CC Diagnosis is by standard immunoassays or hybridisation/amplification
CC reactions. Compounds that modulate expression of the transcripts are
CC potentially useful for treatment of (metastatic) breast cancer, while
CC promoters from the transcripts are used to direct expression, in selected
CC cell types, of e.g. therapeutic genes (also ribozymes or antisense
CC sequences), particularly an antigen-encoding sequence for use in gene or
CC cell-based vaccines. Polypeptides encoded by the transcripts are also
CC useful in vaccines; for diagnosing breast cancer and for raising
CC specific antibodies (Ab). Ab are used to detect the polypeptides or as
CC therapeutic agents. Host cells that produce the polypeptides can be used
CC to expand and isolate populations of educated, antigen-specific immune
CC effecter cells, e.g. cytotoxic T lymphocytes, and these used for
CC adoptive immunotherapy.
XX
SQ Sequence 10 BP; 2 A; 3 C; 1 G; 4 T; 0 other;

Query Match 100.0%; Score 9; DB 21; Length 10;
Best Local Similarity 100.0%; Pred. No. 8.3e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAGGTAAGT 9
|||||||||
Db 9 CAGGTAAGT 1

RESULT 4
AAZ85454/c
ID AAZ85454 standard; DNA; 10 BP.
XX
AC AAZ85454;
XX
DT 07-APR-2000 (first entry)
XX
DE Metastatic breast tumour cell downregulated transcript tag #4688.
XX
KW Human; metastatic breast tumour tissue; breast cancer; tag; primer;
KW non-metastatic breast tumour tissue; gene therapy; anticancer;
KW antimetastatic; vaccine; diagnosis; ss.
XX
OS Homo sapiens.
XX
PN WO9965928-A2.
XX
PD 23-DEC-1999.
XX
PF 18-JUN-1999; 99WO-US13647.
XX
PR 19-JUN-1998; 98US-0089853.
PR 19-JUN-1998; 98US-0089997.
PR 19-JUN-1998; 98US-0090039.
PR 19-JUN-1998; 98US-0090040.
PR 19-JUN-1998; 98US-0090041.
XX
PA (GENZ) GENZYME CORP.
PA (ROBE/) ROBERTS B L.
PA (SHAN/) SHANKARA S.
XX
PI Roberts BL, Shankara S;
XX
DR WPI; 2000-106079/09.
XX
PT Isolated polynucleotides differentially expressed between metastatic
PT and non-metastatic breast cancer cells, useful for diagnosis,
PT prevention and treatment of cancer -
XX
PS Claim 1; Page 184; 219pp; English.
XX
CC AAZ80767 to AAZ83941 represent tags corresponding to distinct
CC transcripts that are preferentially transcribed in the metastatic breast
CC tumour tissue (i.e. are upregulated in metastatic breast tumour cells).
CC AAZ83942 to AAZ86677 represent tags corresponding to distinct transcripts
CC that are preferentially transcribed in the primary or non-metastatic
CC breast tumour tissue (i.e. are downregulated in metastatic breast tumour
CC cells). These transcripts can be used for diagnosis, prognosis,
CC monitoring and treatment of breast cancer, particularly where metastatic.
CC Diagnosis is by standard immunoassays or hybridisation/amplification
CC reactions. Compounds that modulate expression of the transcripts are
CC potentially useful for treatment of (metastatic) breast cancer, while
CC promoters from the transcripts are used to direct expression, in selected
CC cell types, of e.g. therapeutic genes (also ribozymes or antisense
CC sequences), particularly an antigen-encoding sequence for use in gene or
CC cell-based vaccines. Polypeptides encoded by the transcripts are also
CC useful in vaccines; for diagnosing breast cancer and for raising
CC specific antibodies (Ab). Ab are used to detect the polypeptides or as
CC therapeutic agents. Host cells that produce the polypeptides can be used
CC to expand and isolate populations of educated, antigen-specific immune
CC effecter cells, e.g. cytotoxic T lymphocytes, and these used for
CC adoptive immunotherapy.
XX
SQ Sequence 10 BP; 2 A; 3 C; 2 G; 3 T; 0 other;

Query Match 100.0%; Score 9; DB 21; Length 10;
Best Local Similarity 100.0%; Pred. No. 8.3e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAGGTAAGT 9
|||||||||
Db 10 CAGGTAAGT 2

RESULT 5
AAH43531/c
ID AAH43531 standard; DNA; 10 BP.
XX
AC AAH43531;
XX
DT 13-DEC-2001 (first entry)
XX
DE SD sequence.
XX
KW Mouse; heat shock antigen; HSA; human; rat; signal transducer; CD24;
KW fusion protein; inhibition; autoreactive T cell; aTC; primer; PCR;
KW autoimmune disease; multiple sclerosis; rheumatoid arthritis;
KW systemic lupus erythematosus; psoriasis; diabetes; allergy; amplify;
KW transplant rejection; transgenic mouse; polymerase chain reaction; ss.
XX
OS Synthetic.
XX
PN WO200172325-A1.
XX
PD 04-OCT-2001.
XX
PF 29-MAR-2001; 2001WO-US40390.
XX
PR 29-MAR-2000; 2000US-192814P.
XX

PA (OHIS) UNIV OHIO STATE RES FOUND.
XX
PI Liu Y, Zheng P, Bai X;
XX
DR WPI; 2001-611581/70.
XX
PT Inhibiting tissue destruction by autoreactive T cells, useful for
PT treating autoimmune diseases, by administering a heat-shock
PT antigen/CD24 polypeptide or its antibody -
XX
PS Example 2; Page 17; 34pp; English.
XX
CC The sequences given in AAH43525-34 are primers which were used in
CC the production of a fusion gene which comprises a nucleotide
CC sequence encoding the mouse heat shock antigen (HSA) fused to the
CC cDNA sequence of human IgG1 Fc. The resulting fusion protein
CC may be used in the method of the invention for inhibiting
CC destruction of tissue initiated by autoreactive T cells (aTC). The
CC method is especially used to treat subjects suspected of having
CC autoimmune diseases, particularly multiple sclerosis, rheumatoid
CC arthritis, systemic lupus erythematosus, psoriasis, diabetes and
CC allergy, also transplant rejection. Transgenic mice that express
CC human CD24 on their T cells are useful as models for testing drugs
CC for use against autoimmune diseases.
XX
SQ Sequence 10 BP; 2 A; 3 C; 1 G; 4 T; 0 other;

Query Match 100.0%; Score 9; DB 22; Length 10;
Best Local Similarity 100.0%; Pred. No. 8.3e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAGGTAAGT 9
|||||||||
Db 9 CAGGTAAGT 1

RESULT 6
AAH63471/c
ID AAH63471 standard; cDNA; 10 BP.
XX
AC AAH63471;
XX
DT 20-SEP-2001 (first entry)
XX
DE Human ubiquitously expressed transcriptome sequence SEQ ID NO: 311.
XX
KW Human; transcriptome; gene expression pattern; cancer; drug screening;
KW cancer diagnosis; cell specific gene expression; ss.
XX
OS Homo sapiens.
XX
PN WO200138577-A2.
XX
PD 31-MAY-2001.
XX
PF 21-NOV-2000; 2000WO-US31922.
XX
PR 24-NOV-1999; 99US-0448480.
XX
PA (UYJO) UNIV JOHNS HOPKINS.
XX
PI Velculescu VE, Vogelstein B, Kinzler KW;
XX
DR WPI; 2001-367706/38.
XX
PT New isolated polynucleotides, useful for identifying specific cell
PT type, such as cancer cell, comprises transcriptomes expressed in
PT particular cell types -
XX
PS Claim 13; Page 46; 94pp; English.
XX
CC The present invention describes a method of identifying the type of cell
CC in a sample, involving determining which of the sequences
CC AAH63161-AAH64724 is expressed by the cell. The transcriptomes described
CC in the invention are cell-type specific, cancer specific or ubiquitously
CC expressed in humans. They can also be used to screen for drugs, reduce
CC cancer specific gene expression, standardise expression and restore the
CC function of a diseased cell or tissue. The present sequence is one of
CC the transcriptomes described in the exemplification of the invention.
XX
SQ Sequence 10 BP; 2 A; 4 C; 1 G; 3 T; 0 other;

Query Match 100.0%; Score 9; DB 22; Length 10;
Best Local Similarity 100.0%; Pred. No. 8.3e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAGGTAAGT 9
|||||||||
Db 9 CAGGTAAGT 1

RESULT 7
AAH23795
ID AAH23795 standard; DNA; 11 BP.
XX
AC AAH23795;
XX
DT 16-AUG-2001 (first entry)
XX
DE Murine histidine decarboxylase exon/intron boundary #9.
XX
KW Murine; histidine decarboxylase; enzyme; mouse chromosome 2; histamine;
KW ds.
XX
OS Mus sp.
XX
PN WO200132892-A1.
XX
PD 10-MAY-2001.
XX
PF 01-NOV-2000; 2000WO-JP07689.
XX
PR 02-NOV-1999; 99JP-0312559.
XX
PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
PI Ohtsu H;
XX
DR WPI; 2001-308746/32.
XX
PT Polynucleotide encoding histidine decarboxylase located on mouse
PT chromosome 2 for producing histamine defective animal models -
XX
PS Claim 2; Fig 1; 27pp; Japanese.
XX
CC The present invention relates to an isolated and purified polynucleotide
CC located on mouse chromosome 2, encoding histidine decarboxylase,
CC comprising exons 1 to 12 to a total length of approximately 24kb. The
CC present sequence is an exon/intron boundary from the histidine
CC decarboxylase polynucleotide sequence of the present invention.
CC Recombinant vectors containing the polynucleotide with at least one
CC exon substituted by a drug resistance gene, preferably neomycin
CC resistance are also claimed. The polynucleotide is used to produce
CC histamine defective animal models for studying histamine related
CC disorders in humans and to produce treatments for them.
XX
SQ Sequence 11 BP; 4 A; 1 C; 3 G; 3 T; 0 other;

Query Match 100.0%; Score 9; DB 22; Length 11;
Best Local Similarity 100.0%; Pred. No. 8.3e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAGGTAAGT 9

```
          |||||||||
Db      3 caggtaagt 11


RESULT   8
AAZ11271
ID   AAZ11271 standard; DNA; 13 BP.
XX
AC   AAZ11271;
XX
DT   15-NOV-1999  (first entry)
XX
DE   Splice donor site #1 for VP16 gene trap vector.
XX
KW   Splice donor; VP16 gene trap vector; protein-cell interaction; detection;
KW   protein-protein interaction; transcriptional activator domain; ds.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1..4
FT                   /*tag= a
FT                   /label= sticky_end
FT                   /note= "the 5' end of the complementary strand overhangs
FT                          the 3' end of this strand by the sequence
FT                          5'-TCAT-3'"
XX
PN   WO9943848-A1.
XX
PD   02-SEP-1999.
XX
PF   25-FEB-1999;   99WO-CA00173.
XX
PR   25-FEB-1998;   98CA-2224475.
XX
PA   (UYBR-) UNIV BRITISH COLUMBIA.
XX
PI   Jirik FR,  Ong CJ;
XX
DR   WPI; 1999-540605/45.
XX
PT   New protein interaction and transcription factor trap used for
PT   identification of unknown genes encoding transcriptional activator
PT   domains
XX
PS   Example 1; Page 25; 40pp; English.
XX
CC   This sequence represents a splice donor site that can be used in a VP16
CC   gene trap vector used in the method of the invention. The method is for
CC   detecting interaction between an endogenous protein of a cell and a test
CC   protein. The cell contains a first DNA sequence encoding a reporter under
CC   transcriptional control of a transcriptional regulatory element, and a
CC   second DNA sequence that is expressed by the cell and which encodes a
CC   first hybrid protein comprising a first transcriptional regulatory
CC   protein moiety (TRP) selected from a DNA-binding domain that recognises a
CC   binding site on the transcriptional regulatory element controlling
CC   transcriptional of the first DNA sequence and, a transcriptional
CC   activator functional in the cell; and a test protein. The method
CC   comprises: (a) placing into the cell a DNA construct comprising one or
CC   more mRNA splice sites, and a third DNA sequence encoding a second TRP
CC   which, when combined with the first TRP, will reconstitute a
CC   TRP capable of binding to and activating the transcriptional regulatory
CC   element controlling transcription of the first DNA sequence; and
CC   (b) determining whether the reporter is expressed by the cell, as an
CC   indicator of expression of a second hybrid protein comprising the second
CC   TRP and an endogenous protein of the cell capable of interaction with the
CC   test protein. The method is used for the identification and
CC   characterisation of unknown genes according to protein-protein
CC   interactions or for identification of genes encoding transcriptional
CC   activator domains.
XX
SQ   Sequence 13 BP; 4 A; 2 C; 4 G; 3 T; 0 other;

  Query Match             100.0%;  Score 9;  DB 20;  Length 13;
  Best Local Similarity   100.0%;  Pred. No. 8.3e+03;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CAGGTAAGT 9
           |||||||||
Db       5 caggtaagt 13


RESULT   9
AAZ11272
ID   AAZ11272 standard; DNA; 14 BP.
XX
AC   AAZ11272;
XX
DT   15-NOV-1999  (first entry)
XX
DE   Splice donor site #2 for VP16 gene trap vector.
XX
KW   Splice donor; VP16 gene trap vector; protein-cell interaction; detection;
KW   protein-protein interaction; transcriptional activator domain; ds.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1..4
FT                   /*tag= a
FT                   /label= sticky_end
FT                   /note= "the 5' end of the complementary strand overhangs
FT                          the 3' end of this strand by the sequence
FT                          5'-TCAT-3'"
XX
PN   WO9943848-A1.
XX
PD   02-SEP-1999.
XX
PF   25-FEB-1999;   99WO-CA00173.
XX
PR   25-FEB-1998;   98CA-2224475.
XX
PA   (UYBR-) UNIV BRITISH COLUMBIA.
XX
PI   Jirik FR,  Ong CJ;
XX
DR   WPI; 1999-540605/45.
XX
PT   New protein interaction and transcription factor trap used for
PT   identification of unknown genes encoding transcriptional activator
PT   domains
XX
PS   Example 1; Page 25; 40pp; English.
XX
CC   This sequence represents a splice donor site that can be used in a VP16
CC   gene trap vector used in the method of the invention. The method is for
CC   detecting interaction between an endogenous protein of a cell and a test
CC   protein. The cell contains a first DNA sequence encoding a reporter under
CC   transcriptional control of a transcriptional regulatory element, and a
CC   second DNA sequence that is expressed by the cell and which encodes a
CC   first hybrid protein comprising a first transcriptional regulatory
CC   protein moiety (TRP) selected from a DNA-binding domain that recognises a
CC   binding site on the transcriptional regulatory element controlling
CC   transcriptional of the first DNA sequence and, a transcriptional
CC   activator functional in the cell; and a test protein. The method
CC   comprises: (a) placing into the cell a DNA construct comprising one or
CC   more mRNA splice sites, and a third DNA sequence encoding a second TRP
CC   which, when combined with the first TRP, will reconstitute a
CC   TRP capable of binding to and activating the transcriptional regulatory
CC   element controlling transcription of the first DNA sequence; and
CC   (b) determining whether the reporter is expressed by the cell, as an
CC   indicator of expression of a second hybrid protein comprising the second
CC   TRP and an endogenous protein of the cell capable of interaction with the
CC   test protein. The method is used for the identification and
```

```
CC   characterisation of unknown genes according to protein-protein
CC   interactions or for identification of genes encoding transcriptional
CC   activator domains.
XX
SQ   Sequence 14 BP; 4 A; 3 C; 4 G; 3 T; 0 other;


  Query Match             100.0%;  Score 9;  DB 20;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 8.2e+03;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CAGGTAAGT 9
          |||||||||
Db      6 caggtaagt 14


RESULT  10
AAZ40412
ID   AAZ40412 standard; DNA; 15 BP.
XX
AC   AAZ40412;
XX
DT   15-FEB-2000  (first entry)
XX
DE   5' splice site sequence for interferon-alpha plasmid.
XX
KW   Wild type; human; interferon-alpha; plasmid; cytomegalovirus; CMV;
KW   promoter; growth hormone; untranslated region; UTR; mammal; disease;
KW   cancer; intron; ss.
XX
OS   Synthetic.
XX
PN   WO9947678-A2.
XX
PD   23-SEP-1999.
XX
PF   12-MAR-1999;   99WO-US05394.
XX
PR   19-MAR-1998;   98US-0078654.
XX
PA   (GENE-) GENEMEDICINE INC.
XX
PI   Nordstrom J,  Pericle F,  Rolland A,  Ralston R;
XX
DR   WPI; 1999-562116/47.
XX
PT   New plasmids containing an interferon-alpha coding sequence, used for
PT   the treatment of a mammalian condition or disease, particularly cancer
PT   -
XX
PS   Disclosure; Page 31; 137pp; English.
XX
CC   The invention relates to a novel plasmid comprising a cytomegalovirus
CC   (CMV) promoter transcriptionally linked with an interferon alpha
CC   (IFN-alpha) coding sequence, and a growth hormone 3'-untranslated
CC   region (UTR).  Sequences AAZ40412 and AAZ40413 represent synthetic 5' and
CC   3' splice donor and acceptor sites respectively for generating a
CC   synthetic intron to be inserted into the plasmid of the invention. The
CC   plasmids can be used for treating a mammalian condition or disease,
CC   particularly cancer.
XX
SQ   Sequence 15 BP; 3 A; 3 C; 4 G; 5 T; 0 other;


  Query Match             100.0%;  Score 9;  DB 20;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 8.2e+03;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CAGGTAAGT 9
          |||||||||
Db      1 caggtaagt 9


RESULT  11
AAZ11273
ID   AAZ11273 standard; DNA; 15 BP.
XX
AC   AAZ11273;
XX
DT   15-NOV-1999  (first entry)
XX
DE   Splice donor site #3 for VP16 gene trap vector.
XX
KW   Splice donor; VP16 gene trap vector; protein-cell interaction; detection;
KW   protein-protein interaction; transcriptional activator domain; ds.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1..4
FT                   /*tag= a
FT                   /label= sticky_end
FT                   /note= "the 5' end of the complementary strand overhangs
FT                          the 3' end of this strand by the sequence
FT                          5'-TCAT-3'"
XX
PN   WO9943848-A1.
XX
PD   02-SEP-1999.
XX
PF   25-FEB-1999;   99WO-CA00173.
XX
PR   25-FEB-1998;   98CA-2224475.
XX
PA   (UYBR-) UNIV BRITISH COLUMBIA.
XX
PI   Jirik FR,  Ong CJ;
XX
DR   WPI; 1999-540605/45.
XX
PT   New protein interaction and transcription factor trap used for
PT   identification of unknown genes encoding transcriptional activator
PT   domains
XX
PS   Example 1; Page 25; 40pp; English.
XX
CC   This sequence represents a splice donor site that can be used in a VP16
CC   gene trap vector used in the method of the invention. The method is for
CC   detecting interaction between an endogenous protein of a cell and a test
CC   protein. The cell contains a first DNA sequence encoding a reporter under
CC   transcriptional control of a transcriptional regulatory element, and a
CC   second DNA sequence that is expressed by the cell and which encodes a
CC   first hybrid protein comprising a first transcriptional regulatory
CC   protein moiety (TRP) selected from a DNA-binding domain that recognises a
CC   binding site on the transcriptional regulatory element controlling
CC   transcriptional of the first DNA sequence and, a transcriptional
CC   activator functional in the cell; and a test protein. The method
CC   comprises: (a) placing into the cell a DNA construct comprising one or
CC   more mRNA splice sites, and a third DNA sequence encoding a second TRP
CC   which, when combined with the first TRP, will reconstitute a
CC   TRP capable of binding to and activating the transcriptional regulatory
CC   element controlling transcription of the first DNA sequence; and
CC   (b) determining whether the reporter is expressed by the cell, as an
CC   indicator of expression of a second hybrid protein comprising the second
CC   TRP and an endogenous protein of the cell capable of interaction with the
CC   test protein. The method is used for the identification and
CC   characterisation of unknown genes according to protein-protein
CC   interactions or for identification of genes encoding transcriptional
CC   activator domains.
XX
SQ   Sequence 15 BP; 4 A; 4 C; 4 G; 3 T; 0 other;


  Query Match             100.0%;  Score 9;  DB 20;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 8.2e+03;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      1 CAGGTAAGT 9
          |||||||||
Db      7 caggtaagt 15


RESULT  12
AAX71417/c
ID   AAX71417 standard; RNA; 17 BP.
XX
AC   AAX71417;
XX
DT   28-JUL-1999  (first entry)
XX
DE   Human KDR VEGF receptor hammerhead ribozyme substrate #429.
XX
KW   Vascular endothelial growth factor receptor; VEGF receptor; flt-1;
KW   flk-1; KDR; hammerhead ribozyme; hairpin ribozyme; cleavage;
KW   tumour angiogenesis; psoriasis; rheumatoid arthritis; ocular disease;
KW   fms-like tyrosine kinase 1; kinase insert domain containing receptor;
KW   foetal liver kinase 1; ss.
XX
OS   Homo sapiens.
XX
PN   WO9715662-A2.
XX
PD   01-MAY-1997.
XX
PF   25-OCT-1996;   96WO-US17480.
XX
PR   11-JAN-1996;   96US-0584040.
PR   26-OCT-1995;   95US-0005974.
XX
PA   (CHIR ) CHIRON CORP.
PA   (RIBO-) RIBOZYME PHARM INC.
XX
PI   Escobedo J,  McSwiggen J,  Pavco P,  Stinchcomb D;
XX
DR   WPI; 1997-259017/23.
XX
PT   Nucleic acid molecule modulating VEGF receptor(s) gene expression or
PT   mRNA stability  - useful for treating e.g. tumour angiogenesis,
PT   psoriasis, rheumatoid arthritis, etc., in a human patient
XX
PS   Claim 4; Page 110; 218pp; English.
XX
CC   The present invention describes nucleic acid molecules which modulate
CC   the synthesis, expression and/or stability of a mRNA encoding 1 or more
CC   receptors of vascular endothelial growth factor (VEGF). A patient
CC   (preferably human) having a condition associated with the level of the
CC   fms-like tyrosine kinase 1 (flt-1), kinase insert domain containing
CC   receptor (KDR) and/or foetal liver kinase 1 (flk-1) (e.g. tumour
CC   angiogenesis, ocular diseases, psoriasis and rheumatoid arthritis) can
CC   be treated by administering the nucleic acid molecule or the expression
CC   vector to the patient. AAX67275 to AAX75752 represent specific examples
CC   of nucleic acid molecules from the present invention.
XX
SQ   Sequence 17 BP; 3 A; 5 C; 4 G; 5 U; 0 other;


  Query Match            100.0%;  Score 9;  DB 18;  Length 17;
  Best Local Similarity  100.0%;  Pred. No. 8.2e+03;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CAGGTAAGT 9
          |||||||||
Db     14 CAGGTAAGT 6


RESULT  13
AAX71418/c
ID   AAX71418 standard; RNA; 17 BP.
XX
AC   AAX71418;
XX
DT   28-JUL-1999  (first entry)
XX
DE   Human KDR VEGF receptor hammerhead ribozyme substrate #430.
XX
KW   Vascular endothelial growth factor receptor; VEGF receptor; flt-1;
KW   flk-1; KDR; hammerhead ribozyme; hairpin ribozyme; cleavage;
KW   tumour angiogenesis; psoriasis; rheumatoid arthritis; ocular disease;
KW   fms-like tyrosine kinase 1; kinase insert domain containing receptor;
KW   foetal liver kinase 1; ss.
XX
OS   Homo sapiens.
XX
PN   WO9715662-A2.
XX
PD   01-MAY-1997.
XX
PF   25-OCT-1996;   96WO-US17480.
XX
PR   11-JAN-1996;   96US-0584040.
PR   26-OCT-1995;   95US-0005974.
XX
PA   (CHIR ) CHIRON CORP.
PA   (RIBO-) RIBOZYME PHARM INC.
XX
PI   Escobedo J,  McSwiggen J,  Pavco P,  Stinchcomb D;
XX
DR   WPI; 1997-259017/23.
XX
PT   Nucleic acid molecule modulating VEGF receptor(s) gene expression or
PT   mRNA stability  - useful for treating e.g. tumour angiogenesis,
PT   psoriasis, rheumatoid arthritis, etc., in a human patient
XX
PS   Claim 4; Page 110; 218pp; English.
XX
CC   The present invention describes nucleic acid molecules which modulate
CC   the synthesis, expression and/or stability of a mRNA encoding 1 or more
CC   receptors of vascular endothelial growth factor (VEGF). A patient
CC   (preferably human) having a condition associated with the level of the
CC   fms-like tyrosine kinase 1 (flt-1), kinase insert domain containing
CC   receptor (KDR) and/or foetal liver kinase 1 (flk-1) (e.g. tumour
CC   angiogenesis, ocular diseases, psoriasis and rheumatoid arthritis) can
CC   be treated by administering the nucleic acid molecule or the expression
CC   vector to the patient. AAX67275 to AAX75752 represent specific examples
CC   of nucleic acid molecules from the present invention.
XX
SQ   Sequence 17 BP; 4 A; 5 C; 4 G; 4 U; 0 other;


  Query Match            100.0%;  Score 9;  DB 18;  Length 17;
  Best Local Similarity  100.0%;  Pred. No. 8.2e+03;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CAGGTAAGT 9
          |||||||||
Db     13 CAGGTAAGT 5


RESULT  14
AAA36045
ID   AAA36045 standard; DNA; 17 BP.
XX
AC   AAA36045;
XX
DT   26-JUL-2000  (first entry)
XX
DE   Human genomic SNP allele specific oligonucleotide SEQ ID NO:102.
XX
KW   Human; single nucleotide polymorphism; SNP; genotyping; DNA analysis;
KW   allele specific oligonucleotide; ASO; reduced complexity genome; RCG;
KW   genomic classification; identification; DNA fingerprinting;
KW   tumour characterisation; hybridisation; ss.
```

XX
OS Homo sapiens.
XX
PN WO200018960-A2.
XX
PD 06-APR-2000.
XX
PF 24-SEP-1999; 99WO-US22283.
XX
PR 25-SEP-1998; 98US-0101757.
XX
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
XX
PI Landers JE, Jordan B, Housman DE, Charest A;
XX
DR WPI; 2000-293181/25.
XX
PT Detection of single nucleotide polymorphisms in genomes by preparation
PT and analysis of reduced complexity genomes, useful for genotyping,
PT fingerprinting and determining allele frequency of SNPs -
XX
PS Disclosure; Page 56; 111pp; English.
XX
CC A method has been developed for detecting the presence or absence of a
CC single nucleotide polymorphism (SNP) allele in a genomic sample. The
CC method comprises preparing a reduced complexity genome (RCG) from the
CC genomic sample and analysing the RCG for the presence or absence of a
CC SNP allele. The method can be used to characterise a tumour, to generate
CC a genomic pattern for an individual genome or to generate a genomic
CC classification code for a genome. The method can be used to assess
CC whether a subject is at risk for developing a disease or to identify a
CC set of SNP alleles associated with a disease. The method can also be
CC used to perform linkage analysis. AAA35944 to AAA35947 represent
CC sequences used in the exemplification of the present invention. AAA35948
CC to AAA36632 represent nucleotide sequences containing SNPs.
XX
SQ Sequence 17 BP; 5 A; 2 C; 6 G; 4 T; 0 other;

Query Match 100.0%; Score 9; DB 21; Length 17;
Best Local Similarity 100.0%; Pred. No. 8.2e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAGGTAAGT 9
|||||||||
Db 6 caggtaagt 14

RESULT 15
AAC72504/c
ID AAC72504 standard; DNA; 19 BP.
XX
AC AAC72504;
XX
DT 09-FEB-2001 (first entry)
XX
DE Single nucleotide polymorphism PCR primer #1556.
XX
KW Single nucleotide polymorphism; SNP; human; genetic disease;
KW disease susceptibility; cardiovascular system; endocrine system;
KW neurological system; forensic testing; paternity testing; PCR primer; ss.
XX
OS Homo sapiens.
XX
PN WO200058519-A2.
XX
PD 05-OCT-2000.
XX
PF 30-MAR-2000; 2000WO-US08440.
XX
PR 31-MAR-1999; 99US-0127248.
XX
PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
PA (AFFY-) AFFYMETRIX INC.
XX
PI Altshuler D, Cargill M, Daley GQ, Ireland JS, Lander ES;
PI Lipshutz RJ, Patil N, Sklar P;
XX
DR WPI; 2000-611722/58.
XX
PT Nucleic acid selected from one of 106 genes comprising single
PT nucleotide polymorphisms, allele-specific oligonucleotides to the genes
PT are useful for phenotypic correlations, forensics, paternity testing,
PT medicine and genetic analysis -
XX
PS Claim 8; Fig 5; 214pp; English.
XX
CC The present invention is concerned with a number of human single
CC nucleotide polymorphisms (SNPs) which the inventors identified in human
CC genes. These SNPs can be used in disease diagnosis and prediction of an
CC individual's susceptibility to disease, in forensic and paternity testing
CC and in genetic mapping. In particular, the SNPs of the invention can be
CC used to diagnose susceptibility to diseases of the cardiovascular,
CC endocrine and neurological systems, such as coronary artery disease,
CC schizophrenia, cancer, autoimmune diseases, Alzheimer's and Parkinson's
CC diseases.
XX
SQ Sequence 19 BP; 4 A; 4 C; 6 G; 5 T; 0 other;

Query Match 100.0%; Score 9; DB 21; Length 19;
Best Local Similarity 100.0%; Pred. No. 8.2e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAGGTAAGT 9
|||||||||
Db 15 CAGGTAAGT 7

Search completed: July 21, 2002, 09:55:18
Job time: 6379 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2002, 06:28:09 ; Search time 112.48 Seconds
(without alignments)
19.654 Million cell updates/sec

Title: US-09-754-014-10_COPY_1_9
Perfect score: 9
Sequence: 1 CAGGTAAGT 9

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	9	100.0	9	5	PCT-US92-10024-1	Sequence 1, Appli
c	2	9	100.0	14	2	US-08-781-620B-11	Sequence 11, Appl
	3	9	100.0	15	3	US-09-012-366-6	Sequence 6, Appli
c	4	9	100.0	17	4	US-08-584-040-4167	Sequence 4167, Ap
c	5	9	100.0	17	4	US-08-584-040-4168	Sequence 4168, Ap
	6	9	100.0	20	1	US-08-714-626-4	Sequence 4, Appli
	7	9	100.0	20	2	US-08-922-169-4	Sequence 4, Appli
	8	9	100.0	20	4	US-09-226-012-31	Sequence 31, Appl
	9	9	100.0	20	4	US-09-323-743-57	Sequence 57, Appl
	10	9	100.0	20	4	US-09-323-743-58	Sequence 58, Appl
	11	9	100.0	20	4	US-09-323-743-59	Sequence 59, Appl
	12	9	100.0	20	4	US-09-323-743-60	Sequence 60, Appl
	13	9	100.0	20	4	US-09-323-743-61	Sequence 61, Appl
	14	9	100.0	20	4	US-09-323-743-62	Sequence 62, Appl
c	15	9	100.0	20	4	US-09-556-031-10	Sequence 10, Appl
	16	9	100.0	20	5	PCT-US95-04852-4	Sequence 4, Appli
c	17	9	100.0	21	2	US-08-256-426B-274	Sequence 274, App
c	18	9	100.0	23	4	US-09-098-628-50	Sequence 50, Appl
c	19	9	100.0	23	4	US-09-098-628-62	Sequence 62, Appl
	20	9	100.0	28	2	US-08-859-998-1010	Sequence 1010, Ap
	21	9	100.0	28	4	US-09-225-928-1010	Sequence 1010, Ap
	22	9	100.0	30	1	US-08-123-702-34	Sequence 34, Appl
	23	9	100.0	39	1	US-08-257-073-21	Sequence 21, Appl
	24	9	100.0	39	2	US-08-184-009-143	Sequence 143, App
	25	9	100.0	39	2	US-08-458-356-143	Sequence 143, App
	26	9	100.0	39	2	US-08-658-665-152	Sequence 152, App
	27	9	100.0	39	4	US-08-796-101-128	Sequence 128, App
	28	9	100.0	39	4	US-08-460-736-143	Sequence 143, App
	29	9	100.0	39	4	US-09-085-273-152	Sequence 152, App
	30	9	100.0	41	4	US-09-238-356-2	Sequence 2, Appli
c	31	9	100.0	42	2	US-08-792-075-5	Sequence 5, Appli
	32	9	100.0	45	2	US-08-454-557C-54	Sequence 54, Appl
	33	9	100.0	45	2	US-08-340-426D-54	Sequence 54, Appl
	34	9	100.0	45	2	US-08-450-673C-54	Sequence 54, Appl
	35	9	100.0	45	5	PCT-US95-17111A-54	Sequence 54, Appl
c	36	9	100.0	47	1	US-08-147-696E-23	Sequence 23, Appl
c	37	9	100.0	47	1	US-08-147-696E-27	Sequence 27, Appl
c	38	9	100.0	47	1	US-08-484-334-23	Sequence 23, Appl
c	39	9	100.0	47	1	US-08-484-334-27	Sequence 27, Appl
c	40	9	100.0	47	3	US-09-013-092-23	Sequence 23, Appl
c	41	9	100.0	47	3	US-09-013-092-27	Sequence 27, Appl
c	42	9	100.0	47	3	US-09-280-999-23	Sequence 23, Appl
c	43	9	100.0	47	3	US-09-280-999-27	Sequence 27, Appl
c	44	9	100.0	47	4	US-08-952-793-195	Sequence 195, App
c	45	9	100.0	47	5	PCT-US96-09455A-195	Sequence 195, App

ALIGNMENTS

```
RESULT   1
PCT-US92-10024-1
; Sequence 1, Application PC/TUS9210024
;  GENERAL INFORMATION:
;    APPLICANT:  Chang, Tse Wen
;    TITLE OF INVENTION:  ANTI-SENSE OLIGONUCLEOTIDES FOR ISOTYPE-SPECIFIC
;    TITLE OF INVENTION:  SUPPRESSION OF IMMUNOGLOBULIN PRODUCTION
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Tanox Biosystems, Inc.
;      STREET:  10301 Stella Link Rd.
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77025
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  DOS 3.30
;      SOFTWARE:  Wordperfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/10024
;      FILING DATE:  19921118
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/794,395
;      FILING DATE:  11/18/91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mirabel, Eric P.
;      REGISTRATION NUMBER:  31,211
;      REFERENCE/DOCKET NUMBER:  TNX91-6-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (713) 664-2288
;      TELEFAX:  (713) 664-8914
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9 nucleotides
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double-stranded
;      TOPOLOGY:  linear
PCT-US92-10024-1


  Query Match             100.0%;  Score 9;  DB 5;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 2.6e+07;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CAGGTAAGT 9
           |||||||||
Db       1 CAGGTAAGT 9
```

```
RESULT   2
US-08-781-620B-11/c
; Sequence 11, Application US/08781620B
; Patent No. 5994124
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Ribozyme-snRNA chimeric molecules having a
;    TITLE OF INVENTION:  catalytic activity for nuclear RNAs
;    NUMBER OF SEQUENCES:  22
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/781,620B
;      FILING DATE:
;      CLASSIFICATION:  514
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  14 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-781-620B-11


  Query Match             100.0%;  Score 9;  DB 2;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 8.5e+02;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CAGGTAAGT 9
           |||||||||
Db      12 CAGGTAAGT 4


RESULT   3
US-09-012-366-6
; Sequence 6, Application US/09012366
; Patent No. 6034072
;  GENERAL INFORMATION:
;    APPLICANT:  Robert Ralston
;    APPLICANT:  Susanne Muller
;    APPLICANT:  Russ Mumper
;    APPLICANT:  William Munger
;    APPLICANT:  Maria Bruno
;    TITLE OF INVENTION:  IL-2 GENE EXPRESSION AND
;    TITLE OF INVENTION:  DELIVERY SYSTEMS AND USES
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  FastSEQ for Windows 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/012,366
;      FILING DATE:  January 23, 1998
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/039,709
;      FILING DATE:  February 10, 1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkman, Charles S.
;      REGISTRATION NUMBER:  38,077
;      REFERENCE/DOCKET NUMBER:  230/214
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  15 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-012-366-6


  Query Match             100.0%;  Score 9;  DB 3;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 8.5e+02;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CAGGTAAGT 9
           |||||||||
Db       1 CAGGTAAGT 9


RESULT   4
US-08-584-040-4167/c
; Sequence 4167, Application US/08584040
; Patent No. 6346398
;  GENERAL INFORMATION:
;    APPLICANT:  Pavco, Pamela
;    APPLICANT:  McSwiggen, James
;    APPLICANT:  Stinchcomb, Dan T.
;    APPLICANT:  Escobedo, Jaime
;    TITLE OF INVENTION:  METHOD AND REAGENT FOR THE
;    TITLE OF INVENTION:  TREATMENT OF DISEASES OR
;    TITLE OF INVENTION:  CONDITIONS RELATED TO LEVELS
;    TITLE OF INVENTION:  OF VASCULAR ENDOTHELIAL
;    TITLE OF INVENTION:  GROWTH FACTOR
;    NUMBER OF SEQUENCES:  8502
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  Word Perfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/584,040
;      FILING DATE:  January 11, 1996
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/005,974
;      FILING DATE:  October 26, 1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  218/064
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  4167:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  17 base pairs
;      TYPE:  nucleic acid
```

```
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-584-040-4167


  Query Match             100.0%;  Score 9;  DB 4;  Length 17;
  Best Local Similarity   100.0%;  Pred. No. 8.5e+02;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CAGGTAAGT 9
           |||||||||
Db      14 CAGGTAAGT 6


RESULT   5
US-08-584-040-4168/c
; Sequence 4168, Application US/08584040
; Patent No. 6346398
;  GENERAL INFORMATION:
;    APPLICANT:  Pavco, Pamela
;    APPLICANT:  McSwiggen, James
;    APPLICANT:  Stinchcomb, Dan T.
;    APPLICANT:  Escobedo, Jaime
;    TITLE OF INVENTION:  METHOD AND REAGENT FOR THE
;    TITLE OF INVENTION:  TREATMENT OF DISEASES OR
;    TITLE OF INVENTION:  CONDITIONS RELATED TO LEVELS
;    TITLE OF INVENTION:  OF VASCULAR ENDOTHELIAL
;    TITLE OF INVENTION:  GROWTH FACTOR
;    NUMBER OF SEQUENCES:  8502
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  Word Perfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/584,040
;      FILING DATE:  January 11, 1996
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/005,974
;      FILING DATE:  October 26, 1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  218/064
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  4168:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  17 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-584-040-4168


  Query Match             100.0%;  Score 9;  DB 4;  Length 17;
  Best Local Similarity   100.0%;  Pred. No. 8.5e+02;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CAGGTAAGT 9
           |||||||||
Db      13 CAGGTAAGT 5


RESULT   6
US-08-714-626-4
; Sequence 4, Application US/08714626
; Patent No. 5698400
;  GENERAL INFORMATION:
;    APPLICANT:  Cotton, Richard G.H.
;    APPLICANT:  Youil, Rima
;    APPLICANT:  Kemper, Borries W.
;    TITLE OF INVENTION:  Detection of Mutation by
;    TITLE OF INVENTION:  Resolvase Cleavage
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      COMPUTER:  IBM PS/2 Model 50Z or 55SX
;      OPERATING SYSTEM:  MS-DOS (Version 5.0)
;      SOFTWARE:  WordPerfect (Version 5.1)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/714,626
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,530
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Paul T.
;      REGISTRATION NUMBER:  30,162
;      REFERENCE/DOCKET NUMBER:  06253/002001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-5070
;      TELEFAX:  (617) 542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  20
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-714-626-4


  Query Match             100.0%;  Score 9;  DB 1;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 8.5e+02;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CAGGTAAGT 9
           |||||||||
Db       2 CAGGTAAGT 10


RESULT   7
US-08-922-169-4
; Sequence 4, Application US/08922169
; Patent No. 5958692
;  GENERAL INFORMATION:
;    APPLICANT:  Cotton, Richard G.H.
;    APPLICANT:  Youil, Rima
;    APPLICANT:  Kemper, Borries W.
;    TITLE OF INVENTION:  Detection of Mutation by
;    TITLE OF INVENTION:  Resolvase Cleavage
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street
```

```
;       CITY:  Boston
;       STATE:  Massachusetts
;       COUNTRY:  U.S.A.
;       ZIP:  02110-2804
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;       COMPUTER:  IBM PS/2 Model 50Z or 55SX
;       OPERATING SYSTEM:  MS-DOS (Version 5.0)
;       SOFTWARE:  WordPerfect (Version 5.1)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/922,169
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/232,530
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Clark, Paul T.
;       REGISTRATION NUMBER:  30,162
;       REFERENCE/DOCKET NUMBER:  06253/002001
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 542-5070
;       TELEFAX:  (617) 542-8906
;       TELEX:  200154
;   INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  20
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-922-169-4


  Query Match           100.0%;  Score 9;  DB 2;  Length 20;
  Best Local Similarity  100.0%;  Pred. No. 8.5e+02;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CAGGTAAGT 9
          |||||||||
Db      2 CAGGTAAGT 10


RESULT   8
US-09-226-012-31
; Sequence 31, Application US/09226012
; Patent No. 6207383
; GENERAL INFORMATION:
;  APPLICANT: Keating, Mark T.
;  APPLICANT: Splawski, Igor
;  TITLE OF INVENTION: MUTATIONS IN AND GENOMIC STRUCTURE OF HERG - A LONG QT
;  TITLE OF INVENTION: SYNDROME GENE
;  FILE REFERENCE: 2323-136
;  CURRENT APPLICATION NUMBER: US/09/226,012
;  CURRENT FILING DATE: 1999-01-06
;  EARLIER APPLICATION NUMBER: 09/122,847
;  EARLIER FILING DATE: 1998-07-27
;  NUMBER OF SEQ ID NOS: 116
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 31
;   LENGTH: 20
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-226-012-31


  Query Match           100.0%;  Score 9;  DB 4;  Length 20;
  Best Local Similarity  100.0%;  Pred. No. 8.5e+02;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CAGGTAAGT 9
          |||||||||
Db      8 caggtaagt 16


RESULT   9
US-09-323-743-57
; Sequence 57, Application US/09323743
; Patent No. 6214986
; GENERAL INFORMATION:
;  APPLICANT: Bennett, C. Frank
;  APPLICANT: Dean, Nicholas M.
;  APPLICANT: Monia, Brett P.
;  APPLICANT: Nickoloff, Brian J.
;  APPLICANT: Zhang, QingQing
;  TITLE OF INVENTION: Antisense Modulation of bcl-x Expression
;  FILE REFERENCE: ISPH-0368
;  CURRENT APPLICATION NUMBER: US/09/323,743
;  CURRENT FILING DATE: 1999-06-01
;  EARLIER APPLICATION NUMBER: 09/277,020
;  EARLIER FILING DATE: 1998-03-26
;  EARLIER APPLICATION NUMBER: 09/167,921
;  EARLIER FILING DATE: 1998-10-07
;  NUMBER OF SEQ ID NOS: 66
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 57
;   LENGTH: 20
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic
US-09-323-743-57


  Query Match           100.0%;  Score 9;  DB 4;  Length 20;
  Best Local Similarity  100.0%;  Pred. No. 8.5e+02;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CAGGTAAGT 9
          |||||||||
Db      2 caggtaagt 10


RESULT  10
US-09-323-743-58
; Sequence 58, Application US/09323743
; Patent No. 6214986
; GENERAL INFORMATION:
;  APPLICANT: Bennett, C. Frank
;  APPLICANT: Dean, Nicholas M.
;  APPLICANT: Monia, Brett P.
;  APPLICANT: Nickoloff, Brian J.
;  APPLICANT: Zhang, QingQing
;  TITLE OF INVENTION: Antisense Modulation of bcl-x Expression
;  FILE REFERENCE: ISPH-0368
;  CURRENT APPLICATION NUMBER: US/09/323,743
;  CURRENT FILING DATE: 1999-06-01
;  EARLIER APPLICATION NUMBER: 09/277,020
;  EARLIER FILING DATE: 1998-03-26
;  EARLIER APPLICATION NUMBER: 09/167,921
;  EARLIER FILING DATE: 1998-10-07
;  NUMBER OF SEQ ID NOS: 66
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 58
;   LENGTH: 20
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic
US-09-323-743-58


  Query Match           100.0%;  Score 9;  DB 4;  Length 20;
  Best Local Similarity  100.0%;  Pred. No. 8.5e+02;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CAGGTAAGT 9
```

```
          |||||||||
Db      4 caggtaagt 12


RESULT  11
US-09-323-743-59
; Sequence 59, Application US/09323743
; Patent No. 6214986
; GENERAL INFORMATION:
;  APPLICANT: Bennett, C. Frank
;  APPLICANT: Dean, Nicholas M.
;  APPLICANT: Monia, Brett P.
;  APPLICANT: Nickoloff, Brian J.
;  APPLICANT: Zhang, QingQing
;  TITLE OF INVENTION: Antisense Modulation of bcl-x Expression
;  FILE REFERENCE: ISPH-0368
;  CURRENT APPLICATION NUMBER: US/09/323,743
;  CURRENT FILING DATE: 1999-06-01
;  EARLIER APPLICATION NUMBER: 09/277,020
;  EARLIER FILING DATE: 1998-03-26
;  EARLIER APPLICATION NUMBER: 09/167,921
;  EARLIER FILING DATE: 1998-10-07
;  NUMBER OF SEQ ID NOS: 66
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 59
;   LENGTH: 20
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic
US-09-323-743-59


  Query Match           100.0%;  Score 9;  DB 4;  Length 20;
  Best Local Similarity  100.0%;  Pred. No. 8.5e+02;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CAGGTAAGT 9
          |||||||||
Db      6 caggtaagt 14


RESULT  12
US-09-323-743-60
; Sequence 60, Application US/09323743
; Patent No. 6214986
; GENERAL INFORMATION:
;  APPLICANT: Bennett, C. Frank
;  APPLICANT: Dean, Nicholas M.
;  APPLICANT: Monia, Brett P.
;  APPLICANT: Nickoloff, Brian J.
;  APPLICANT: Zhang, QingQing
;  TITLE OF INVENTION: Antisense Modulation of bcl-x Expression
;  FILE REFERENCE: ISPH-0368
;  CURRENT APPLICATION NUMBER: US/09/323,743
;  CURRENT FILING DATE: 1999-06-01
;  EARLIER APPLICATION NUMBER: 09/277,020
;  EARLIER FILING DATE: 1998-03-26
;  EARLIER APPLICATION NUMBER: 09/167,921
;  EARLIER FILING DATE: 1998-10-07
;  NUMBER OF SEQ ID NOS: 66
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 60
;   LENGTH: 20
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic
US-09-323-743-60


  Query Match           100.0%;  Score 9;  DB 4;  Length 20;
  Best Local Similarity  100.0%;  Pred. No. 8.5e+02;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CAGGTAAGT 9
          |||||||||
Db      8 caggtaagt 16


RESULT  13
US-09-323-743-61
; Sequence 61, Application US/09323743
; Patent No. 6214986
; GENERAL INFORMATION:
;  APPLICANT: Bennett, C. Frank
;  APPLICANT: Dean, Nicholas M.
;  APPLICANT: Monia, Brett P.
;  APPLICANT: Nickoloff, Brian J.
;  APPLICANT: Zhang, QingQing
;  TITLE OF INVENTION: Antisense Modulation of bcl-x Expression
;  FILE REFERENCE: ISPH-0368
;  CURRENT APPLICATION NUMBER: US/09/323,743
;  CURRENT FILING DATE: 1999-06-01
;  EARLIER APPLICATION NUMBER: 09/277,020
;  EARLIER FILING DATE: 1998-03-26
;  EARLIER APPLICATION NUMBER: 09/167,921
;  EARLIER FILING DATE: 1998-10-07
;  NUMBER OF SEQ ID NOS: 66
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 61
;   LENGTH: 20
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic
US-09-323-743-61


  Query Match           100.0%;  Score 9;  DB 4;  Length 20;
  Best Local Similarity  100.0%;  Pred. No. 8.5e+02;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CAGGTAAGT 9
          |||||||||
Db     10 caggtaagt 18


RESULT  14
US-09-323-743-62
; Sequence 62, Application US/09323743
; Patent No. 6214986
; GENERAL INFORMATION:
;  APPLICANT: Bennett, C. Frank
;  APPLICANT: Dean, Nicholas M.
;  APPLICANT: Monia, Brett P.
;  APPLICANT: Nickoloff, Brian J.
;  APPLICANT: Zhang, QingQing
;  TITLE OF INVENTION: Antisense Modulation of bcl-x Expression
;  FILE REFERENCE: ISPH-0368
;  CURRENT APPLICATION NUMBER: US/09/323,743
;  CURRENT FILING DATE: 1999-06-01
;  EARLIER APPLICATION NUMBER: 09/277,020
;  EARLIER FILING DATE: 1998-03-26
;  EARLIER APPLICATION NUMBER: 09/167,921
;  EARLIER FILING DATE: 1998-10-07
;  NUMBER OF SEQ ID NOS: 66
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 62
;   LENGTH: 20
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic
US-09-323-743-62
```

```
  Query Match             100.0%;  Score 9;  DB 4;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 8.5e+02;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CAGGTAAGT 9
           |||||||||
Db      12 caggtaagt 20


RESULT  15
US-09-556-031-10/c
; Sequence 10, Application US/09556031
; Patent No. 6350868
; GENERAL INFORMATION:
;  APPLICANT:  Weston, Brent W.
;  APPLICANT:  Hiller, Kara B.
;  TITLE OF INVENTION:  Antisense Fucosyltransferase Sequences and Methods of
;  TITLE OF INVENTION:  Use Thereof
;  FILE REFERENCE:  Weston and Hiller
;  CURRENT APPLICATION NUMBER: US/09/556,031
;  CURRENT FILING DATE:  2000-04-20
;  PRIOR APPLICATION NUMBER:  60/131,068
;  PRIOR FILING DATE:  1999-04-26
;  NUMBER OF SEQ ID NOS:  24
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 10
;   LENGTH: 20
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:antisense
;   OTHER INFORMATION: oligonucleotide
US-09-556-031-10


  Query Match             100.0%;  Score 9;  DB 4;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 8.5e+02;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CAGGTAAGT 9
           |||||||||
Db      13 CAGGTAAGT 5


Search completed: July 21, 2002, 09:47:18
Job time: 11949 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2002, 06:18:04 ; Search time 3274.61 Seconds
(without alignments)
37.095 Million cell updates/sec

Title: US-09-754-014-10_COPY_1_9
Perfect score: 9
Sequence: 1 CAGGTAAGT 9

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	9	100.0	24	12	AZ478673	AZ478673 1M0298J20
	2	9	100.0	25	12	AZ868376	AZ868376 2M0180N05
c	3	9	100.0	29	12	AZ596214	AZ596214 1M0409A21
c	4	9	100.0	34	10	H40658	H40658 yn79b07.r1
	5	9	100.0	37	9	AA961266	AA961266 on96a05.s
c	6	9	100.0	38	12	AZ461182	AZ461182 1M0266M21
	7	9	100.0	41	9	AW059896	AW059896 30_comp15
c	8	9	100.0	50	9	AU102848	AU102848 AU102848
c	9	9	100.0	50	9	AU102849	AU102849 AU102849
c	10	9	100.0	50	9	AU102850	AU102850 AU102850
	11	9	100.0	50	9	AU103835	AU103835 AU103835
c	12	9	100.0	51	10	BI824201	BI824201 603040569
c	13	9	100.0	51	10	D19969	D19969 HUMGS00934
c	14	9	100.0	51	10	BE978061	BE978061 bs73c08.y
c	15	9	100.0	54	12	AZ346887	AZ346887 1M0082P17
	16	9	100.0	55	12	AZ919892	AZ919892 1006017B0
	17	9	100.0	60	12	B04047	B04047 cSRL-23c11-
	18	9	100.0	61	10	BG271723	BG271723 nai52g06.
	19	9	100.0	61	12	BH626214	BH626214 1007112B0
	20	9	100.0	63	12	AQ026164	AQ026164 1(3)01029
c	21	9	100.0	65	12	AZ773615	AZ773615 2M0001J07
	22	9	100.0	66	12	B04318	B04318 cSRL-30b4-u
c	23	9	100.0	67	12	CNS03MSX	AL251178 Tetraodon
c	24	9	100.0	68	9	AA138566	AA138566 mr08b12.r
c	25	9	100.0	70	9	AA028449	AA028449 mi19f10.r
	26	9	100.0	70	9	AI014447	AI014447 os34d01.s
c	27	9	100.0	71	12	AZ317589	AZ317589 1M0036L22
c	28	9	100.0	71	12	AZ623786	AZ623786 1M0461A19
c	29	9	100.0	71	12	AG025331	AG025331 Oryza sat
	30	9	100.0	74	9	AW611631	AW611631 hg93e10.x
	31	9	100.0	76	12	BH234085	BH234085 1006177D0
c	32	9	100.0	77	9	AV833086	AV833086 AV833086
	33	9	100.0	77	12	AZ495680	AZ495680 1M0331J11
c	34	9	100.0	79	9	AA490024	AA490024 ab02b06.s
	35	9	100.0	79	12	FR0024865	AL017705 F.rubripe
	36	9	100.0	81	12	BH234141	BH234141 1006177G0
	37	9	100.0	82	9	AI311824	AI311824 qo95e07.x
	38	9	100.0	82	9	AI353568	AI353568 zeh0683.s
c	39	9	100.0	83	12	AZ776655	AZ776655 2M0010B03
	40	9	100.0	84	9	AA835987	AA835987 oc81g05.s
c	41	9	100.0	85	9	AI789299	AI789299 uk53e01.y
	42	9	100.0	86	9	AI353306	AI353306 zeh0311.s
c	43	9	100.0	86	10	W63933	W63933 md77a02.r1
	44	9	100.0	87	9	AA231789	AA231789 RZ500.R c
	45	9	100.0	89	10	N88337	N88337 K3123F Huma

ALIGNMENTS

RESULT 1
AZ478673/c
LOCUS AZ478673 24 bp DNA linear GSS 04-OCT-2000
DEFINITION 1M0298J20R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0298J20 R, DNA sequence.
ACCESSION AZ478673
VERSION AZ478673.1 GI:10637794
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 24)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0298 row: J column: 20
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 24.
FEATURES Location/Qualifiers
source 1. .24
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0298J20"
/clone_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 6 a 8 c 3 g 7 t
ORIGIN

Query Match 100.0%; Score 9; DB 12; Length 24;
Best Local Similarity 100.0%; Pred. No. 6.5e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAGGTAAGT 9
|||||||||
Db 14 CAGGTAAGT 6

RESULT 2
AZ868376
LOCUS AZ868376 25 bp DNA linear GSS 21-FEB-2001
DEFINITION 2M0180N05F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0180N05 F, DNA sequence.
ACCESSION AZ868376
VERSION AZ868376.1 GI:13071628
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 25)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0180 row: N column: 05
Seq primer: CGTTGTAAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 25.
FEATURES Location/Qualifiers
source 1. .25
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0180N05"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 8 a 3 c 8 g 6 t
ORIGIN

Query Match 100.0%; Score 9; DB 12; Length 25;
Best Local Similarity 100.0%; Pred. No. 6.5e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAGGTAAGT 9
|||||||||
Db 4 CAGGTAAGT 12

RESULT 3
AZ596214/c
LOCUS AZ596214 29 bp DNA linear GSS 13-DEC-2000
DEFINITION 1M0409A21F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0409A21 F, DNA sequence.
ACCESSION AZ596214
VERSION AZ596214.1 GI:11718404
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 29)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0409 row: A column: 21
Seq primer: CGTTGTAAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 29.
FEATURES Location/Qualifiers
source 1. .29
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"

```
                /clone="UUGC1M0409A21"
                /clone_lib="Mouse 10kb plasmid UUGC1M library"
                /sex="Male"
                /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                /note="Vector: PWD42nv; Purified genomic DNA from M.
                musculus C57BL/6J (male) was obtained  from the Jackson
                Laboratory Mouse DNA Resource
                (http://www.jax.org/resources/documents/dnares/). The DNA
                was hydrodynamically sheared by repeated passage through a
                0.005 inch orifice at constant velocity. The sheared DNA
                was blunt end-repaired with T4 DNA polymerase and T4
                polynucleotide kinase. Adaptor oligonucleotides were
                ligated to the blunt ends in high  molar excess. The
                adaptored DNA was purified and size-selected for a 9.5 to
                10.5 kb range using preparative agarose gel
                electrophoresis. Vector DNA was prepared from a derivative
                of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                inducible derivative of plasmid R1. The vector was ligated
                with adaptors complementary to the insert adaptors and
                purified. The sheared, adaptored mouse DNA was annealed to
                adaptored vector DNA, and transformed into
                chemically-competent E. coli XL10-Gold (Stratagene) cells
                and selected for ampicillin resistance."
BASE COUNT        6 a     11 c      6 g      6 t
ORIGIN


  Query Match             100.0%;  Score 9;  DB 12;  Length 29;
  Best Local Similarity   100.0%;  Pred. No. 6.6e+04;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CAGGTAAGT 9
           |||||||||
Db      22 CAGGTAAGT 14


RESULT   4
H40658/c
LOCUS       H40658                    34 bp    mRNA    linear   EST 31-JUL-1995
DEFINITION  yn79b07.r1 Soares adult brain N2b5HB55Y Homo sapiens cDNA clone
            IMAGE:174613 5' similar to SP:SSRB_CANFA P23438 SIGNAL SEQUENCE
            RECEPTOR BETA SUBUNIT PRECURSOR ;, mRNA sequence.
ACCESSION   H40658
VERSION     H40658.1  GI:916710
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 34)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 781
            High quality sequence starts: 1
            High quality sequence stops: 1
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Trace considered overall poor quality
            Possible reversed clone: similarity on wrong strand
            Insert Length: 781   Std Error: 0.00
            Seq primer: M13Rev
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .34
                     /organism="Homo sapiens"
                     /db_xref="GDB:3836648"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:174613"
                     /clone_lib="Soares adult brain N2b5HB55Y"
                     /sex="Male"
                     /dev_stage="55-year old"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: brain; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGCTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 53. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. The adult brain RNA was
                     provided by Dr. Donald H. Gilden. Tissue was acquired
                     17-18 hours after death which occurred in consequence of a
                     ruptured aortic aneurysm. RNA was prepared from a pool of
                     tissues representing the following areas of the brain:
                     frontal, parietal, temporal and occipital cortex from the
                     left and right hemispheres, subcortical white matter,
                     basal ganglia, thalamus, cerebellum, midbrain, pons and
                     medulla."
BASE COUNT        9 a     12 c      5 g      8 t
ORIGIN


  Query Match             100.0%;  Score 9;  DB 10;  Length 34;
  Best Local Similarity   100.0%;  Pred. No. 6.6e+04;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CAGGTAAGT 9
           |||||||||
Db      31 CAGGTAAGT 23


RESULT   5
AA961266
LOCUS       AA961266                  37 bp    mRNA    linear   EST 23-JUN-1998
DEFINITION  on96a05.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:1564496 3' similar to SW:YGF4_YEAST P53173 HYPOTHETICAL 15.9
            KD PROTEIN IN OLE1-DUP1 INTERGENIC REGION. ;, mRNA sequence.
ACCESSION   AA961266
VERSION     AA961266.1  GI:3127283
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 37)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Trace considered overall poor quality
            Insert Length: 681   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .37
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1564496"
                     /clone_lib="Soares_NFL_T_GBC_S1"
```

/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NbHL19W, testis NHT, and B-cell NCI_CGAP_GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 14 a 4 c 11 g 8 t
ORIGIN

Query Match 100.0%; Score 9; DB 9; Length 37;
Best Local Similarity 100.0%; Pred. No. 6.7e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAGGTAAGT 9
|||||||||
Db 16 CAGGTAAGT 24

RESULT 6
AZ461182/c
LOCUS AZ461182 38 bp DNA linear GSS 04-OCT-2000
DEFINITION 1M0266M21R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0266M21 R, DNA sequence.
ACCESSION AZ461182
VERSION AZ461182.1 GI:10619307
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 38)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0266 row: M column: 21
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 38.
FEATURES Location/Qualifiers
source 1. .38
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0266M21"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 11 a 13 c 2 g 12 t
ORIGIN

Query Match 100.0%; Score 9; DB 12; Length 38;
Best Local Similarity 100.0%; Pred. No. 6.7e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAGGTAAGT 9
|||||||||
Db 17 CAGGTAAGT 9

RESULT 7
AW059896
LOCUS AW059896 41 bp mRNA linear EST 23-AUG-2000
DEFINITION 30_comp15-s30 UPC15 Homo sapiens cDNA similar to CYTOCHROME C OXIDASE POLYPEPTIDE VIB, mRNA sequence.
ACCESSION AW059896
VERSION AW059896.1 GI:6652218
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 41)
AUTHORS Brenner,S., Williams,S.R., Vermass,E.H., Storck,T., Moon,K., McCollum,C., Mao,J.I., Kirchner,J.J., Eletr,S., DuBridge,R.B., Burcham,T. and Albrecht,G.
TITLE In vitro cloning of complex mixtures of DNA on microbeads: Physical separation of differentially expressed cDNAs
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (4), 1665-1670 (2000)
MEDLINE 20144098
COMMENT Contact: Burcham TS
LYNX Therapeutics, Inc.
25861 Industrial Blvd., Hayward, CA 94545, USA
Tel: 510 670 9338
Fax: 510 670 9302
Email: timb@lynxgen.com
Sequence obtained from LYNX Therapeutics Megasort technology.
High quality sequence stop: 41.
FEATURES Location/Qualifiers
source 1. .41
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="UPC15"
/cell_type="monocyctic leukemia"
/cell_line="THP-1 (TIB-202)"
/note="Vector: pCR2.1; Cloning of PCR products from micro-beads carrying 3' end of up-regulated cDNA. THP-1 cells induced with 100 nM PMA in DMSO. "
BASE COUNT 12 a 5 c 9 g 15 t
ORIGIN

Query Match 100.0%; Score 9; DB 9; Length 41;
Best Local Similarity 100.0%; Pred. No. 6.7e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 CAGGTAAGT 9
          |||||||||
Db     19 CAGGTAAGT 27


RESULT   8
AU102848/c
LOCUS       AU102848                  50 bp    mRNA    linear   EST 30-AUG-2001
DEFINITION  AU102848 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
            COLF4743, mRNA sequence.
ACCESSION   AU102848
VERSION     AU102848.1  GI:13552369
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 50)
  AUTHORS   Suzuki,Y., Taira,H., Tsunoda,T., Mizushima-Sugano,J., Sese,J., Hata
            ,H., Ota,T., Isogai,T., Tanaka,T., Morishita,S., Okubo,K., Sakaki
            ,Y., Nakamura,Y., Suyama,A. and Sugano,S.
  TITLE     Diverse transcriptional initiation revealed by fine, large-scale
            mapping of mRNA start sites
  JOURNAL   EMBO Rep. 2 (5), 388-393 (2001)
  MEDLINE   21270072
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp
            Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
            ,S. Construction and characterization of a full length-enriched and
            a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).
FEATURES             Location/Qualifiers
     source          1. .50
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="COLF4743"
                     /clone_lib="Sugano Homo sapiens cDNA library"
BASE COUNT       13 a     11 c     21 g      5 t
ORIGIN


  Query Match             100.0%;  Score 9;  DB 9;  Length 50;
  Best Local Similarity   100.0%;  Pred. No. 6.9e+04;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CAGGTAAGT 9
          |||||||||
Db     10 CAGGTAAGT 2


RESULT   9
AU102849/c
LOCUS       AU102849                  50 bp    mRNA    linear   EST 30-AUG-2001
DEFINITION  AU102849 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
            HSI00737, mRNA sequence.
ACCESSION   AU102849
VERSION     AU102849.1  GI:13552370
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 50)
  AUTHORS   Suzuki,Y., Taira,H., Tsunoda,T., Mizushima-Sugano,J., Sese,J., Hata
            ,H., Ota,T., Isogai,T., Tanaka,T., Morishita,S., Okubo,K., Sakaki
            ,Y., Nakamura,Y., Suyama,A. and Sugano,S.
  TITLE     Diverse transcriptional initiation revealed by fine, large-scale
            mapping of mRNA start sites
  JOURNAL   EMBO Rep. 2 (5), 388-393 (2001)
  MEDLINE   21270072
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp
            Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
            ,S. Construction and characterization of a full length-enriched and
            a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).
FEATURES             Location/Qualifiers
     source          1. .50
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HSI00737"
                     /clone_lib="Sugano Homo sapiens cDNA library"
BASE COUNT       11 a     12 c     19 g      8 t
ORIGIN


  Query Match             100.0%;  Score 9;  DB 9;  Length 50;
  Best Local Similarity   100.0%;  Pred. No. 6.9e+04;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CAGGTAAGT 9
          |||||||||
Db     14 CAGGTAAGT 6


RESULT  10
AU102850/c
LOCUS       AU102850                  50 bp    mRNA    linear   EST 30-AUG-2001
DEFINITION  AU102850 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
            LNG03935, mRNA sequence.
ACCESSION   AU102850
VERSION     AU102850.1  GI:13552371
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 50)
  AUTHORS   Suzuki,Y., Taira,H., Tsunoda,T., Mizushima-Sugano,J., Sese,J., Hata
            ,H., Ota,T., Isogai,T., Tanaka,T., Morishita,S., Okubo,K., Sakaki
            ,Y., Nakamura,Y., Suyama,A. and Sugano,S.
  TITLE     Diverse transcriptional initiation revealed by fine, large-scale
            mapping of mRNA start sites
  JOURNAL   EMBO Rep. 2 (5), 388-393 (2001)
  MEDLINE   21270072
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp
            Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
            ,S. Construction and characterization of a full length-enriched and
            a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).
FEATURES             Location/Qualifiers
     source          1. .50
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="LNG03935"
                     /clone_lib="Sugano Homo sapiens cDNA library"
BASE COUNT       11 a     13 c     19 g      7 t
ORIGIN


  Query Match             100.0%;  Score 9;  DB 9;  Length 50;
  Best Local Similarity   100.0%;  Pred. No. 6.9e+04;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CAGGTAAGT 9
          |||||||||
Db     14 CAGGTAAGT 6
```

```
RESULT  11
AU103835
LOCUS       AU103835                 50 bp    mRNA    linear   EST 30-AUG-2001
DEFINITION  AU103835 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
            HEP06842, mRNA sequence.
ACCESSION   AU103835
VERSION     AU103835.1  GI:13553356
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 50)
  AUTHORS   Suzuki,Y., Taira,H., Tsunoda,T., Mizushima-Sugano,J., Sese,J., Hata
            ,H., Ota,T., Isogai,T., Tanaka,T., Morishita,S., Okubo,K., Sakaki
            ,Y., Nakamura,Y., Suyama,A. and Sugano,S.
  TITLE     Diverse transcriptional initiation revealed by fine, large-scale
            mapping of mRNA start sites
  JOURNAL   EMBO Rep. 2 (5), 388-393 (2001)
  MEDLINE   21270072
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp
            Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
            ,S. Construction and characterization of a full length-enriched and
            a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).
FEATURES             Location/Qualifiers
     source          1. .50
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HEP06842"
                     /clone_lib="Sugano Homo sapiens cDNA library"
BASE COUNT       12 a     14 c     12 g     12 t
ORIGIN


  Query Match             100.0%;  Score 9;  DB 9;  Length 50;
  Best Local Similarity   100.0%;  Pred. No. 6.9e+04;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CAGGTAAGT 9
          |||||||||
Db     19 CAGGTAAGT 27


RESULT  12
BI824201/c
LOCUS       BI824201                 51 bp    mRNA    linear   EST 04-OCT-2001
DEFINITION  603040569F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5181425 5',
            mRNA sequence.
ACCESSION   BI824201
VERSION     BI824201.1  GI:15935751
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 51)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11452  row: j  column: 18
            High quality sequence stop: 51.
FEATURES             Location/Qualifiers
     source          1. .51
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5181425"
                     /clone_lib="NIH_MGC_115"
                     /lab_host="DH10B"
                     /note="Organ: pooled brain, lung, testis; Vector:
                     pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
                     source anonymous pool of 6 male brains, age range 23-27; 1
                     male lung, age 27; and 1 male testis, age 69.  Library is
                     oligo-dT primed and directionally cloned (EcoRV site is
                     destroyed upon  cloning).  Average insert size 1.8 kb,
                     insert size range 1-3 kb. Library is normalized and
                     enriched for full-length clones and was  constructed by C.
                     Gruber (Invitrogen).  Research Genetics tracking  code
                     021. Note: this is a NIH_MGC Library."
BASE COUNT       11 a     12 c     18 g     10 t
ORIGIN


  Query Match             100.0%;  Score 9;  DB 10;  Length 51;
  Best Local Similarity   100.0%;  Pred. No. 6.9e+04;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CAGGTAAGT 9
          |||||||||
Db     25 CAGGTAAGT 17


RESULT  13
D19969/c
LOCUS       D19969                   51 bp    mRNA    linear   EST 30-JUL-1996
DEFINITION  HUMGS00934 Human promyelocyte Homo sapiens cDNA clone mm0994 3',
            mRNA sequence.
ACCESSION   D19969
VERSION     D19969.1  GI:500866
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 51)
  AUTHORS   Okubo,K., Fukushima,A., Yoshii,J., Niiyama,T., Kojima,Y., Yoshinari
            ,H., Arimoto,J. and Matsubara,K.
  TITLE     Gene expression of human promyelocytic cell line HL60 before and
            after induction of differentiation. A new application of 3'directed
            cDNA sequencing
  JOURNAL   Unpublished (1993)
COMMENT     Contact: Okubo,K., Fukushima,A., Yoshii,J., Niiyama,T., Kojima,Y.,
            Yoshinari,H., Arimoto,J. and Matsubara,K.
            Institute for Molecular and Cellular Biology
            Osaka University
            3-1 Yamada-oka,Suita,Osaka 565,Japan.
FEATURES             Location/Qualifiers
     source          1. .51
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="mm0994"
                     /clone_lib="Human promyelocyte"
                     /note="Female, adult, cell_line = HL60, cell_type =
                     promyelocyte. "
BASE COUNT       22 a      9 c      7 g     13 t
ORIGIN


  Query Match             100.0%;  Score 9;  DB 10;  Length 51;
  Best Local Similarity   100.0%;  Pred. No. 6.9e+04;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CAGGTAAGT 9
          |||||||||
```

```
Db     23 CAGGTAAGT 15


RESULT  14
BE978061/c
LOCUS       BE978061                 51 bp    mRNA    linear   EST 04-OCT-2000
DEFINITION  bs73c08.y1 Drosophila melanogaster adult testis library Drosophila
            melanogaster cDNA clone bs73c08 5', mRNA sequence.
ACCESSION   BE978061
VERSION     BE978061.1  GI:10609159
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 51)
  AUTHORS   Andrews, J., Bouffard, G. and Oliver, B.
  TITLE     Drosophila melanogaster testis expressed sequence tags
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Brian Oliver
            Laboratory of Cellular and Developmental Biology
            NIDDK, National Institutes of Health
            6 Center Drive MSC 2715, Bldg 6, Rm B1-13, Bethesda, MD 20892 USA
            Fax: (301) 496 5239
            Email: oliver@helix.nih.gov,
            http://www.niddk.nih.gov/intram/people/boliver.htm
            Tissue isolation and library construction performed at the National
            Institute of Diabetes and Digestive and Kidney Diseases, NIH (see
            http://www.niddk.nih.gov/intram/people/boliver.htm). DNA sequencing
            and analyses performed by National Institutes of Health Intramural
            Sequencing Center (NISC; see http://www.nisc.nih.gov).
            Plate: 73  row: c  column: 08
            Seq primer: M13RP1 reverse primer (ABI).
FEATURES             Location/Qualifiers
     source          1. .51
                     /organism="Drosophila melanogaster"
                     /strain="y[*] w[67c1]/Y"
                     /db_xref="taxon:7227"
                     /clone="bs73c08"
                     /clone_lib="Drosophila melanogaster adult testis library"
                     /sex="male"
                     /dev_stage="1-5 day adult"
                     /lab_host="SOLR (Stratagene)"
                     /note="Organ: testis; Vector: pBlueScript SK (Stratagene);
                     Site_1: EcoR I; Site_2: Xho I; Testes dissected from 1-5
                     day adult y[*] w[67c1]/Y males raised at 25oC.  RNA
                     isolated using Trizol (Life Technologies) and a single
                     round of Poly(A)+ selection using Oligotex (Qiagen).  cDNA
                     library constructed using Stratagene ZAP-cDNA syntesis
                     kit.  Oligo dT-primed, size fractionated ~1-6 kb, and
                     directionally cloned at EcoRI and XhoI in Uni-ZAP XR.
                     Following a single round of amplification pBlueScript SK
                     phagemids were mass excised.  A distribution channel for
                     clones is being sought, but not currently available.
                     Requests for clones cannot be honored."
BASE COUNT       12 a     11 c     12 g     16 t
ORIGIN


  Query Match           100.0%;  Score 9;  DB 10;  Length 51;
  Best Local Similarity  100.0%;  Pred. No. 6.9e+04;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CAGGTAAGT 9
          |||||||||
Db     41 CAGGTAAGT 33


RESULT  15
AZ346887/c
LOCUS       AZ346887                 54 bp    DNA     linear   GSS 29-SEP-2000
DEFINITION  1M0082P17F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0082P17 F, DNA sequence.
ACCESSION   AZ346887
VERSION     AZ346887.1  GI:10426124
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 54)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0082  row: P  column: 17
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 54.
FEATURES             Location/Qualifiers
     source          1. .54
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0082P17"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       11 a     13 c     14 g     16 t
ORIGIN


  Query Match           100.0%;  Score 9;  DB 12;  Length 54;
  Best Local Similarity  100.0%;  Pred. No. 6.9e+04;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CAGGTAAGT 9
          |||||||||
Db     42 CAGGTAAGT 34



Search completed: July 21, 2002, 09:11:03
Job time: 10379 sec
```